THE UNIVERSITY
of EDINBURGH

In vivo and *in vitro* characterization and application of tyrosine recombinases for metabolic engineering

Wei Liu

Thesis submitted for the degree of

Doctor of Philosophy

The University of Edinburgh

2017

Abstract

Site-specific recombinases are a family of DNA-modifying enzymes that can recognize short specific DNA sequences and drive recombination between them to rearrange DNA fragments which results in excision, integration or inversion. Here I describe *in vivo* application of the recombinases for chassis optimization for heterologous pathway expression in synthetic yeast and *in vitro* application of recombinases for gene expression optimization.

For *in vivo* application, two recombination systems, Cre/loxP and Dre/rox, were developed for driving orthogonal co-SCRaMbLE of normal synthetic chromosomes and the tRNA Neochromosome. The functions of two recombinases were engineered to be activated and controlled by two hormone molecules, β-estradiol and RU486. In addition, a SCRaMbLE-in device was designed to integrate a pathway of interest into a synthetic chromosome in yeast while driving the normal SCRaMbLE process at the same time for chassis level optimization by the Cre/loxP recombination system. For *in vitro* application, three recombinases, Cre, Dre and VCre, were explored to integrate promoters into a pathway of interest for gene expression level diversification in metabolic engineering.

To attempt broader application of the recombinases for metabolic engineering, a side project in metabolic engineering was also involved in my PhD study. As part of a collaborative Synthetic Natural Product (SynNP) project applying synthetic biology to discover and design new antibiotics against tuberculosis and other infectious diseases, a YeastFab compatible assembly method was designed for large pathway construction and tested for heterologous expression of the RiPPs exemplar pathway of nocathiacin I in *S. cerevisiae*.

Lay Summary

Site-specific recombinases are enzymes that can recognize short specific DNA sequences and drive recombination between them to achieve DNA excision, integration or inversion. An international consortium of scientists has been working on re-writing the genome of baker's yeast by putting recombination sites across the yeast genome. The recombinases can then be used to generate random modifications to accelerate the evolution of the species. I developed and engineered a set of recombinases to further engineer the yeast for metabolic engineering applications, screening for evolved yeast strains with higher production of products of interest. The activation of the function of the recombinases was designed to be controlled by hormone chemicals. The recombinases were also extracted and purified so the DNA can be easily and quickly recombined outside the cell. This *in vitro* recombinase toolkit was also applied to metabolic engineering. I also worked on a collaborative Synthetic Natural Product (SynNP) project applying synthetic biology to discover and design new antibiotics against tuberculosis and other infectious diseases.

Declaration

I declare that the thesis has been composed by myself and the work is my own. The work has not been submitted for any other degree or professional qualification. Any included publications are my own work, unless otherwise stated.

Wei Liu
August 2017

Acknowledgements

First, I would like to express sincere gratitude to my supervisor Professor Patrick Yizhi Cai for giving me the opportunity to do a PhD in his lab at the University of Edinburgh, for his guidance and mentorship on the projects during my PhD study, for his efforts and motivation to build us a nice working environment and bring all the fantastic people together to work as a great team. I also want to express my sincere thanks to my supervisor Professor Chris French for his suggestions and helps in solving many research problems during my study, and I much appreciated his dedicated revisions and corrections of my PhD thesis. I also thank my research committee, Professor Susan Rosser and Professor Peter Swain for insightful suggestions on my 10-month report, 2nd year poster, 3rd year presentation and other training sessions. I want to thank my studentship sponsor, school of biological science and the Autodesk company, for joint financial support for my PhD and funding supports from Bill & Melinda Gates Foundation.

Many thanks to our collaborators Nhan Pham and Irene Pérez pi from Professor Manfred Auer's lab for the help with the HPLC analysis of violacein; many thanks to Laura Tuck from Dr. Jon Marles-Wright for the help with the chromatography purification of the recombinases; many thanks to Yun Wang and Longying Liu from Chantal Shen's group at the Beijing Genomics Institute for the next generation sequencing and LC-MS analysis of carotenoids; many thanks to Professor Mike Tyers and Professor Gerry Wright for the guidance in the collaborative SynNP project.

I am also grateful to work with my lab members and appreciate their support and discussions on lab techniques. I thank Chantal and Roy for sharing the *SynII* strain and tRNA array plasmids for recombinase function characterization; I thank Dr. Dariusz Abramczyk for sharing the synthetic PCR tags for SCRaMbLE test assays; I thank Dr. Eva Garcia-Ruiz and Jamie Auxillos for their supports on the YeastFab promoter library construction and discussions on the SynNP projects. Special thanks to lab staff Aileen and previous lab staffs Alba and Isaac for the support with media support, lab management and IT support, which makes our life so much easier in the lab.

Finally, my deepest love goes to my parents and my boyfriend. Without my parents' openness, trust, understanding and full support, I won't be able to pursue the PhD education in UK to broaden my horizon in the world famous university. I am so touched by all the efforts that my boyfriend, Yang Liu, made to join me in Edinburgh for a PhD study. His scientific opinions, encouragements and sometimes criticisms are beneficial to me and keep me motivated to meet all the challenges. I wish him the best and very success in his future academic career and want to tell him,

Thank you and I love you!

Table of contents

List of Figures

List of Tables

List of abbreviations

5-FOA	5-Fluoroorotic Acid
Amp	Ampicillin
APS	Ammonium persulfate
BGC	Biosynthetic gene cluster
BSA	Bovine serum albumin
BSG	Biosynthetic gene
CDS	Coding sequence
DEPC	Diethyl pyrocarbonate
DMSO	Dimethyl sulphoxide
EBD	Estrogen binding domain
EDTA	Ethylenediaminetetraacetic acid
EoI	Element of Interest
FACS	Fluorescence activated cell sorting
FCC	Frozen competent cell
GFP	Green fluorescent protein
GoI	Gene of Interest
HPLC	High-performance liquid chromatography
HS DNA	Herring sperm DNA
IDT	Integrated DNA Technologies
IPTG	Isopropyl thiogalactopyranoside
Kan	Kanamycin
LB	Lysogeny broth
LC-MS	Liquid chromatography–mass spectrometry
LE	Left element
MIC	Minimal inhibition concentration
NADPH	Nicotinamide adenine dinucleotide phosphate
NC	Negative control
NGS	Next generation sequencing

NP	Natural product
NRPS	Non-ribosomal peptide synthase
OPM	Outer primer mix
ORF	Open reading frame
PBD	Progesterone binding domain
PBS	Phosphate buffered saline
PC	Positive control
PEG	Polyethylene glycol
PKS	Polyketide synthetase
POT	Promoter-open reading frame-terminator
PTM	Posttranslational modification
RE	Right element
RiPP	Ribosomally synthesized and post-translationally modified peptide
RT-PCR	Reverse transcription PCR
SC	Synthetic complete
SCRaMbLE (SCR)	Synthetic chromosome rearrangement and modification by loxP mediated evolution
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SGD	*Saccharomyces* Genome Database
TAE	Tris-acetate-EDTA
TB	Tuberculosis
TEMED	Tetramethylethylenediamine
TFA	trifluoroacetic acid
TPM	Templateless primer mix
TU	Transcription unit
WT	Wild type
YAC	Yeast artificial chromosome
YPD	Yeast Peptone Dextrose

Chapter 1 Introduction and literature review

1.1 Synthetic Biology

1.1.1 Overview and brief history of synthetic biology

Synthetic biology has been developing rapidly in recent decades as a nascent interdisciplinary engineering subject (Figure 1.1). Based on the principles of modern cell and molecular biology, it also incorporates and absorbs design principles from the fields of engineering to mix and match biological functional elements as basic building blocks to construct coordinated devices just as other engineered facilities or systems are constructed.

'Prehistoric' synthetic biology. The studies of the *lac* operon regulatory mechanism in *E. coli* by Francois Jacob and Jacques Monod in 1961 are taken as roots for the origin of synthetic biology (Cameron, Bashor, & Collins, 2014). However, manipulating biological elements is not as easy as traditional well developed engineering methods that are based on the laws of physics and chemistry. The fundamental elements to be engineered in biology are mainly biomacromolecules like DNA, RNA and protein and according to the genetic central dogma, DNA is the final target to be dealt with. With the discovery and development of DNA restriction enzymes, DNA ligase and the invention of PCR during the 1960s to 1980s, the predecessor of synthetic biology, genetic engineering, became a pioneering subject of biotechnology. Genetic engineering stood out from traditional biology, yet its ability was still limited both because of its limitations on DNA cloning techniques, number of genes that can be modified and also insufficient understanding of biological systems. One of the most important techniques in the 1990s, DNA sequencing, brought genetic research into the era of "omics" and "high-throughput", greatly increasing the number of DNA engineering elements available. Meanwhile, researchers focused on using a top-down method to study genetic networks by breaking them down to subsystems or interrupting one or several of the network components to investigate the function of these elements in the whole system. The improvement of all these techniques in molecular biology and deeper knowledge in

systems biology laid foundations for synthetic biology.

Establishment of synthetic biology. Deep understanding of the properties of individual elements, taking advantage of the characterized components and building a new system with them from the bottom up are the core features of synthetic biology. Before 1995, Harley and Lucy compared the difference and similarity between genetic networks and electronic circuits and analyzed how signal timing and time delay in the network could generate different dynamic patterns with mathematical modeling (McAdams & Shapiro, 1995). Their work was very important to systems biology by applying insights from electronic engineering to bacterial cell networks, and this was also the time when the term “genetic circuit” was born. This concept had been widely used in synthetic biology thereafter. At the beginning of the new century, the first synthetic biological circuit, a toggle switch designed by J. Collins and his colleagues, was published (Gardner, Cantor, & Collins, 2000). It achieved control of the cell to switch between two states stably with the change of input signals in *E. coli*. The toggle switch was achieved by the interaction of two promoter-repressor systems, with IPTG and high temperature being the two inputs that can relieve the repression effect and lead to one-way change of state. In the same year, the repressilator was designed with a third system added to the two promoter-repressor pairs, resulting in periodic oscillation of output fluorescence (Elowitz & Leibler, 2000). These works are two examples of synthetic biological systems. In addition to these two leading designs, the first wave of synthetic biology also includes other gene regulatory networks like switches fulfilled by other activator or repressor systems or in mammalian cells (Atkinson, Savageau, Myers, & Ninfa, 2003; Kramer et al., 2004), cascades (Hooshangi, Thiberge, & Weiss, 2005), combinatorial gene networks (Guet, Elowitz, Hsing, & Leibler, 2002), intercellular communication devices based on quorum-sensing systems (Basu, Mehreja, Thiberge, Chen, & Weiss, 2004), time-delay circuits (W. Weber et al., 2007), oscillators (Stricker et al., 2008; Tigges, Marquez-Lago, Stelling, & Fussenegger, 2009) and natural product pathway circuits (Martin, Pitera, Withers, Newman, & Keasling, 2003), etc.

Though simple, the design of the toggle switch and repressilator illustrated the

universal “part-circuit-device” structure of a synthetic biological system. Parts are the basic genetic components, such as inducible promoters and their corresponding repressor proteins; then the combination of a certain promotor and repressor protein form a circuit or module, which is further combined with others to constitute the final device that is expected to show a certain behavior or fulfill a desired function. Synthetic biology requires rational design of a system with knowledge of the characteristics of basic components. Therefore, part characterization and behavior prediction with proper mathematical modeling are the key to constructing a robust and reliable synthetic device. As in other engineering fields, the system designed does not always work as expected and the “Design-Build-Test-Learn” cycle also applies to synthetic biology. Quantitative measurement and mathematical modelling will help to evaluate cell behavior more accurately and make it more predictive for future designs (Isaacs, Hasty, Cantor, & Collins, 2003). Cellular context is a key confounding factor that can affect the function of the circuit (Blake, Kaern, Cantor, & Collins, 2003). In additional to noise, if the function of the part is not orthogonal to the cell, it can result in malfunction of the part or even toxicity to the host. With specific engineering of parts to target host organisms, the noises and side effects can be reduced or removed and part function be optimized by various directed evolution method (Yokobayashi, Weiss, & Arnold, 2002).

Expansion of synthetic biology. At the beginning of synthetic biology, the work was to some extent confined to a few leading groups including those of James Collins, Ron Weiss, George Church, Jay Keasling, Craig Venter, etc. As a young subject, it needed broader spread and more contributions to grow faster and become more influential. The traditional way to assemble biological parts is to use different restriction enzymes to cut and ligate the basic elements. Since different groups have different preference for the enzymes, it makes it inconvenient and time consuming to customize the DNA cloning every time in a new lab or for a new project. To facilitate the sharing of the parts available and enhance communication within the synthetic biology community, part standardization is necessary. In 2003, Tom Knight developed the “BioBrick” standard which uses vectors with the cutting sites of EcoRI, SpeI, XbaI

and PstI for easy assembly of two parts, and further developed more vectors for circuit assembly (Shetty, Endy, & Knight, 2008). Following the lead of these research groups, a teaching course on synthetic biology was opened at MIT for the younger generation to play with the DNA BioBricks (Smolke, 2009). The course later became a competition between several universities, which later became the first iGEM (international Genetically Engineered Machine competition) to design functional genetic devices using the standardized parts. The format of the competition each year is a great way to teach the concepts of synthetic biology, to inspire innovation and stimulate creativity among undergraduate students. Teams participating in the competition share and contribute to the library of interchangeable standardized parts, which became the Registry of Standard Biological Parts (http://parts.igem.org/Main_Page). The accumulation of the parts benefits subsequent iGEMers and even outside groups, allowing them to build more diverse and complex circuits. Beside the official wiki of iGEM, labs can also use an open wiki platform, OpenWetWare, to communicate methodology and protocols for assembly and part characterization. The scale and influence of the competition has been growing extremely fast. In the 2016 iGEM, over 5600 participants across 42 countries took part in the competition, and many former iGEMers have continued their work in synthetic biology for doctoral study and started their own labs of synthetic biology. Brand new ideas and breakthroughs are coming out every year with deeper and more thorough description of the systems. It is the combined intelligence in the community that has promoted and vitalized synthetic biology.

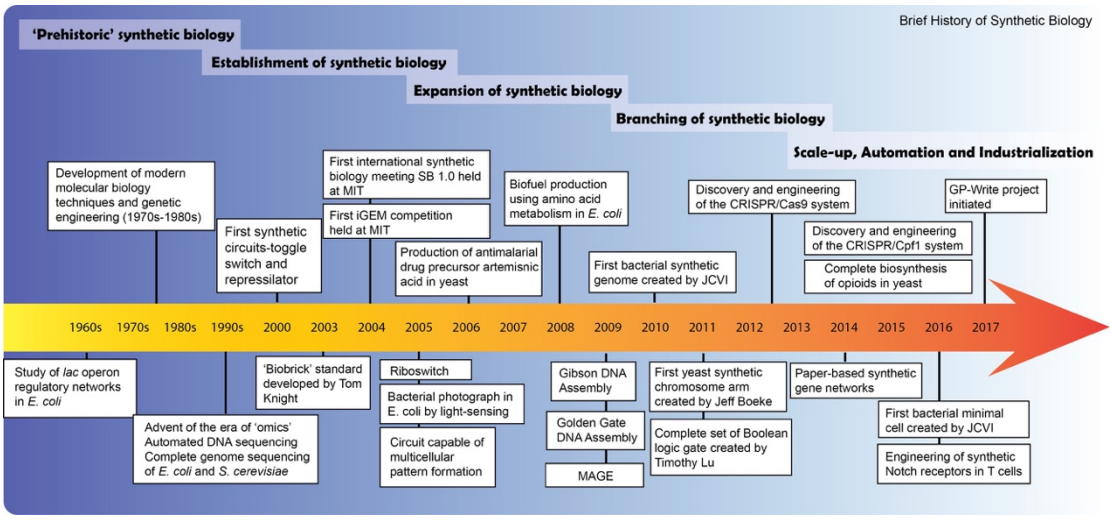


Figure 1.1| The development of synthetic biology (Figure modified from J. Collin's review (Cameron et al., 2014))

Branching of synthetic biology. Since synthetic biology is a young engineering discipline with a broad choice of circuits to design and problems to solve, many sub-directions and fields of application have emerged with its expansion. In general, it can be mainly categorized into fundamental research and application oriented research. Most of the early designs focused on fundamental study, including genetic switches like toggle switches (Atkinson et al., 2003; Gardner et al., 2000; Green, Silver, Collins, & Yin, 2014; Isaacs et al., 2003; Kramer et al., 2004), computational devices like logic gates (Daniel, Rubens, Sarpeshkar, & Lu, 2013; Miyamoto, Razavi, DeRose, & Inoue, 2013; Siuti, Yazbek, & Lu, 2013; L. Wang et al., 2015), DNA recorder devices (Farzadfard & Lu, 2014), bio-computers (Baumgardner et al., 2009; Haynes et al., 2008), RNA synthetic biology like riboswitches (Bayer & Smolke, 2005; Chappell, Watters, Takahashi, & Lucks, 2015; Isaacs, Dwyer, & Collins, 2006; Mandal & Breaker, 2004; Serganov & Patel, 2007), biosensors (Daszczuk et al., 2014; Lapique & Benenson, 2014; Samodelov et al., 2016; Slomovic, Pardee, & Collins, 2015; Watstein, McNerney, & Styczynski, 2015), DNA origami (Han et al., 2011; Ke et al., 2009; Maune et al., 2010), micro-compartments and protein scaffolds (Agapakis, Boyle, & Silver, 2012; Bashor, 2008; Chessher, Breitling, & Takano, 2015), intercellular communication (Basu et al., 2004; Brenner, You, & Arnold, 2008; You, Cox, Weiss, & Arnold, 2004), programmed cell pattern formation (Basu, Gerchman, Collins, Arnold,

& Weiss, 2005), biosafety devices (Y. Cai et al., 2015; Wright, Stan, & Ellis, 2013), etc. Beside the tools above, one of the hottest topics in synthetic biology is the engineering of genome editing tools like ZFNs, TALEN and CRISPR/Cas9 (Gaj, Gersbach, & Barbas, 2013; Hsu, Lander, & Zhang, 2014; Wood et al., 2011). The invention and development of the CRISPR system enables specific targeting in the genome and facilitates further genetic engineering of theoretically any organism with a sequenced genome (Mali et al., 2013). In return, this top down method of studying specific gene function in the genome provides more resources to be standardized for future design. Based on the fundamental research, many applications of synthetic biology have been attempted to solve real life problems in fields like environmental protection, healthcare and energy sustainability (Khalil & Collins, 2010). Combined with metabolic engineering, synthetic biology approaches have been used to improve biofuel production (Atsumi, Hanai, & Liao, 2008; d'Espaux, Mendez-Perez, Li, & Keasling, 2015), to produce bio-materials (Chessher et al., 2015; R. Langer & Tirrell, 2004; Wagner, Sprenger, Rebmann, & Weber, 2016), pharmaceuticals and other high-value chemicals (Chang & Keasling, 2006; Galanie, Thodey, Trenchard, Filsinger Interrante, & Smolke, 2015; Keasling, 2012; Paddon & Keasling, 2014a; Thodey, Galanie, & Smolke, 2014). There are also various other applications ranging from bio-printing (Levskaya et al., 2005), biosensors for pollution detection or disease diagnosis (Courbet, Renard, & Molina, 2016; Joshi, Wang, Montgomery, Elfick, & French, 2009; Kaur, Kumar, Babu, & Mittal, 2015; Pardee et al., 2014; Schmidt & Cho, 2015), to gene therapy and other health related applications (Abil, Xiong, & Zhao, 2015; Agustin-Pavon & Isalan, 2014; Anderson, Clarke, Arkin, & Voigt, 2006; Dobrin, Saxena, & Fussenegger, 2016; Ruder, Lu, & Collins, 2011; W. Weber & Fussenegger, 2011). As to the scale of the synthetic circuit, the trend is moving from simple circuits like single transcriptional units to modular circuits to more complicated circuits with multi-layer feedback, and finally genome level synthesis. This trend is made possible by the continued development of diverse DNA assembly methods that enable faster multi-part assembly and decreasing cost of DNA sequencing and synthesis. As to the chassis and platform, the host organism is spreading from prokaryotes, like the

starting model organism *E. coli*, to *Streptomyces* (Medema, Breitling, Bovenberg, & Takano, 2011; Medema, Breitling, & Takano, 2011*)*, *Cyanobacteria* (Berla et al., 2013; Huang, Camsund, Lindblad, & Heidorn, 2010; B. Wang, Wang, Zhang, & Meldrum, 2012), and higher to eukaryotes, like the model yeast strain *S.cerevisiae* (Cantone et al., 2009; Krivoruchko, Siewers, & Nielsen, 2011; Siddiqui, Thodey, Trenchard, & Smolke, 2012), plants (Baltes & Voytas, 2015; Liu, Yuan, & Stewart, 2013; Sarrion-Perdigones et al., 2013) and recently mammalian cells (Auslander & Fussenegger, 2013; Kis, Pereira, Homma, Pedrigi, & Krams, 2015; Lienert, Lohmueller, Garg, & Silver, 2014), and lower to minimal cell and cell-free systems (Forster & Church, 2007; Hodgman & Jewett, 2012; Jewett, Calhoun, Voloshin, Wuu, & Swartz, 2008; Kuruma, Stano, Ueda, & Luisi, 2009). Since our knowledge about the even the simplest organisms is still limited, It is worth building up different cell factories and taking advantage of their special features for different applications. With deeper studies on cell-free synthetic biology, we can isolate different cellular components and study their biochemical characteristics in a simpler system and achieve rational design for difficult circuits and devices.

Scale-up, automation and industrialization. With all the advances in fundamental and application related research, synthetic biology is also entering the stage of commercial transformation. Typically, a synthetic biology project includes the process of design, construction and function characterization. The process of circuit design is the section that requires the most human intelligence. Accurate construction of the circuits is also important but with a well-established protocol for DNA assembly, it is less difficult. However, unlike writing a computer language, the construction section is more time and manpower consuming, especially as the number of circuits needed is increasing all the time. If a circuit is not working as expected, the troubleshooting process always includes repeated construction from the start. Since high-throughput and repetition are the two features most required in the industrialization of synthetic biology, new technologies for automation must be pushed forward to free scientists and engineers from labor work and allow them to focus on creation and innovation. As the most powerful tool of the 20th century,

the computer has assisted and benefited almost every field of human life. The same is true for synthetic biology. Besides the diverse databases, which perform static storage of different DNA, RNA and protein sequences, more interactive tools to facilitate circuit design and construction are being developed. First of all, to standardize the description of the DNA element, Synthetic Biology Open Language (SBOL) was established (Galdzicki et al., 2014; Quinn et al., 2015). Many computer-aided design (CAD) tools are available for various purposes in circuit design, including SnapGene, Benchling, j5, GenoCAD, BIOFAB studio, etc (Carothers, Goler, Juminaga, & Keasling, 2011; Chandran, Bergmann, & Sauro, 2009; Chandran, Bergmann, Sauro, & Densmore, 2011; Hillson, Rosengarten, & Keasling, 2012). As for computer-aided manufacturing (CAM), many pieces of robotic and automation equipment, like the Fluent® and Freedom EVO® series from TECAN, and Automated Nucleic Acid Purification from ThermoFisher, have been invented and put in use. These machines can be used separately to fulfill partial automation in the lab, whereas the final goal of manufacturing in synthetic biology is to achieve a complete automated pipeline in the lab or factory. Therefore, many DNA foundries, including iBioFAB in Illinois, "The London Foundry" at SynbiCITE and Edinburgh Genome Foundry, have been founded to compile all different automation facilities and build integrative workflows for DNA synthesis and assembly. As to automation for characterization, some microfluidics platforms have been used to integrate assembly and characterization (Linshiz et al., 2016), yet this is still difficult on a large scale since the methods of functional analysis are too diverse and not easy to standardize. Future work on simplifying and standardizing characterization is necessary. Besides the DNA synthesis companies like IDT, Gen9 and Twist Bioscience, many other biotech start-ups aiming for different application areas are sprouting. Though synthetic biology is still at a very early stage of industrialization, it has a very promising market and has great potential to become an indispensable commercial field for human society.

1.1.2 Design strategies and characterization

In order to manipulate genetic elements and design a system with predictable behavior, the function and inherent characteristics have to be known for each

element. On top of the knowledge of natural genetic parts, there are some basic principles to follow for circuit design to fit the parts into target devices or systems.

Design strategies. Synthetic biology is an engineering discipline, and some of the design principles can be passed on from traditional engineering subjects. In 2005, Drew Endy discussed the theory and practice of engineering with scholars from engineering fields and proposed the "Foundations for engineering biology", which includes standardization, decoupling and abstraction (Endy, 2005). In the article, he proposed that these three aspects and related technologies developed upon these ideas are the foundation of synthetic biology. This is not only true for the subject macroscopically, but also applies to the design of specific circuits. Standardization has been executed through the history of synthetic biology in both assembly methods and part characterization, which facilitates the re-use of standardized parts for new creations. Decoupling means that the artificially designed device or system can work independently and orthogonally without interacting with the biological context, and its function remains when transferred to a different chassis or platform. It also means to decompose the whole complicated system into individual simplified modules to design, fabricate and characterize. Abstraction refers to extracting the key information of each level of the design (part, device and system) from the complex biological context and simplifying the system to a level such that a mathematical model can be used to evaluate the dynamic behavior. Therefore, the synthetic biological engineer should be able to use standardized language to design the systems modularly like a programmer and think through the method to characterize and quantify the function even at the design stage. Beside the three classical strategies, the inherent characteristics of diverse parts and systems should be considered to flexibly adjust the design and optimize the target functions (Andrianantoandro, Basu, Karig, & Weiss, 2006).

Characterization and quantification. Synthetic biology has spanned diverse fields and served many different goals. For different applications, there are various methods for part characterization. Though diverse, there are common basic rules or universal reporters applicable for the different designs. Most synthetic biological

devices involve transcription and translation (TX-TL) processes and the measurement of transcription or translation efficiency is fundamental to evaluate the performance. In the early years of synthetic biology, two quantification characteristics, PoPS (Polymerase Per Second) and RIPS (Ribosomes Per Second) were proposed to describe the rate of TX-TL. For real experimental measurement, reporters are needed to evaluate the parameters of a device or system. Fluorescent tags or luciferase are the most commonly used reporters for the quantification of protein translation and localization. Transcriptional quantification methods have been developed in recent years, such as RT-qPCR. Beside quantification, other aspects of characterization, like orthogonality of input signals and specificity of feedback signals, are also very important. The quantification methods for TX-TL process are now well standardized, while the fact that the features of biological parts are so diverse and complicated requires more progress to be made in imaging and single enzyme activity for metabolic pathways. Some progress has been achieved by the invention and application of flow cytometry and microfluidics for single cell level characterization, and characterization by cell-free systems for catalytic proteins in metabolic engineering (Dudley, Karim, & Jewett, 2015).

Performance assessment. When characterization of the elementary parts and individual modules is completed, the system as a whole should be further evaluated on its stability, burden on host, robustness, noise and sensitivity/response speed. Stability means whether the whole system can exist and function in the host organism stably with little interference from outside and without generating toxicity. If not toxic, the exogenously introduced system can still generate a burden on the host. Tom Ellis and his colleagues used GFP to indicate such effects in *E. coli* (Ceroni, Algar, Stan, & Ellis, 2015). Robustness means the system's ability to resist perturbation, and for synthetic biology specifically it refers to how well the function of a synthetic device or system can remain in different cell contexts or chassis. Noise, also known as cell heterogeneity or stochastic gene expression, is another variable which needs to be considered (Ciechonska, Grob, & Isalan, 2016). For biosensors, the response speed is a key parameter for performance assessment. The target

machinery of the input signals makes a big difference to the sensitivity. Some devices require fast response to the input and some need a relatively longer response time to fulfill certain feedback functions like the lagging time for oscillators.

In sum, when designing a new synthetic circuit, device or system, it is always good to include how to standardize the part for easy assembly, what characterization method to choose and the availability of corresponding tools, read through the performance of relevant previously characterized systems, and engineer one's own work based on this to have a higher chance of designing new functional devices.

1.1.3 Part standardization and DNA assembly

DNA assembly methods. After circuit design is completed, the next key thing before fulfilling the function of a device is to choose the fittest and fastest DNA assembly method for the construction. There are many different assembly methods, either *in vitro* or *in vivo*. But from the perspective of assembly mechanism, basically they can be categorized to three families: endonuclease-mediated assembly, long-overlap-based assembly and site-specific recombination-based assembly (Casini, Storch, Baldwin, & Ellis, 2015). On top of the normal cut and ligation by normal restriction endonucleases (RE) and ligases, some REs with the same digestion overhangs were chosen to standardize bioparts for idempotent assembly of two fragments with a vector, like BioBrick and BglBrick (Anderson et al., 2010; Shetty et al., 2008). These methods were good inventions in the early days but their assembly ability is also limited, with limited number of parts that can be assembled and remaining scars after ligation. Later, a group of assembly methods based on the special family of restriction endonuclease, the type IIs RE, were developed and widely used, including Golden Gate, Moclo, and GoldenBraid (Engler, Gruetzner, Kandzia, & Marillonnet, 2009; Sarrion-Perdigones et al., 2013; E. Weber, Engler, Gruetzner, Werner, & Marillonnet, 2011). Since the overhangs of the cutting site can be artificially designed and the recognition sites have directionality, there are many advantages of this type of assembly, for example multi-part cloning with designed order is achievable, the target circuit is scarless, and it is flexible to switch parts from a standardized library to achieve device engineering. However, it requires that no cutting site exists inside

the component DNA fragments, and point mutation is therefore sometimes required before assembly. For the long-overlap-based assembly methods, they all require homologous sequences generally longer than 20bp but the mechanism for different methods varies, and they can be either *in vitro* or *in vivo*. Gibson assembly is the most popular and widely used method of this kind. Though it needs the combined use of three enzymes and is relatively expensive, it does not require the absence of specific restriction sites, and can assemble multiple large fragments (Gibson et al., 2009). Besides *in vitro* DNA repair, the assembly can also be achieved *in vivo* through homologous recombination by natural DNA repair machinery in live microbes like *E. coli*, *Bacillus subtilis* and *Saccharomyces cerevisiae*. The invention of circular polymerase extension cloning (CPEC) combined both *in vitro* PCR amplification and *in vivo* DNA repair. It is much cheaper than Gibson assembly while being equally fast and efficient (Quan & Tian, 2009). Recently, the *in vivo* recombination machinery of *E. coli* has been extracted for *in vitro* use in methods like SLiCE and PaperClip, which offer low cost for scaling up without compromising the speed of construction (Trubitsyna, Michlewski, Cai, Elfick, & French, 2014; Y. Zhang, Werling, & Edelmann, 2012). Site-specific recombination based assembly is younger than the others and is fulfilled by the site-specific integrases (Colloms et al., 2014a; L. Zhang, Zhao, & Ding, 2011). Currently, they are mainly used for metabolic pathway engineering in eukaryotes. Though some long scars are inevitable, with the help of an accessory recombination directionality factor, the integration process can be reversed, which facilitates flexibly switching parts in the circuit to achieve circuit diversity.

Choice and combined usage of DNA assembly methods. As to how to choose the right method, it depends on the characteristics of the ingredient DNA and the size of the final circuit. Normally Golden Gate assembly is the first to consider since multiple parts can be assembled by one-pot ligation. However, if there are many forbidden sites, or the total length of a circuit is too long, other methods like Gibson or CPEC and *in vivo* recombination assembly are good options. Sometimes one method cannot meet the need to fully assemble a circuit, and multi-round hierarchical construction can be used to accelerate the speed and at the same time ensure the

quality of construction. There are many examples of hierarchical construction: YeastFab assembly is a specific Golden Gate assembly method that alternatively uses BsaI and BsmBI for pathway construction (Y. Guo et al., 2015); DNA Assembler combines *in vitro* overlap extension and *in vivo* yeast recombination for large pathway construction (Shao, Zhao, & Zhao, 2009). In the international project to synthesize the genome of *S.cerevisiae*, multiple methods were used at different scales for the construction, including overlap-extension PCR, Gibson assembly, yeast *in vivo* assembly and auxotrophic marker swapping (Dymond et al., 2011; Jovicevic, Blount, & Ellis, 2014). Other methods like MODAL and PaperClip, which use standardized linkers or adaptors, are also good complementary methods that can be combined with Gibson Assembly, CPEC, yeast *in vivo* assembly or ligase cycled reaction assembly (LCR) for circuit construction (Casini et al., 2014; Trubitsyna et al., 2014).

1.2 Synthetic genomes and genome editing

1.2.1 Synthetic genome overview

One branching field of synthetic biology is the synthesis of the whole genome of a natural organism. From the point of assembly scale, it is of the highest level; from the point of rational design, it is also of the most difficulty. The first synthetic genome was that of a prokaryote *Mycoplasma mycoides* with the length of 1.08 Mbp, created by J. Craig Venter Institute (JCVI) (Gibson et al., 2010). The chemically synthesized genome successfully transformed *M. capricolum* to *M. mycoides*. In the version 1.0 synthesized genome, most genetic information was maintained and only some watermark sequences were introduced together with several designed gene deletions. Though not many changes were made, this proved that an artificially synthesized genome can be maintained stably in a natural organism and function similarly to the natural one, which is an important milestone in synthetic biology, proving technical feasibility of genome synthesis. However, merely copying nature's design is not the goal of synthetic biology. The next step is to rationally add functional design in new target genomes. The efforts include designing minimal genomes, rearranging genes modularly by function, and adding heterologous or fully artificial

sequences to the new genomes. JCVI continued their work on designing minimized genome based on the JCV-syn1.0 strain and recently they published their newest progress on synthesizing the JCV-syn 3.0 genome with length and gene number reduced to 531kb and 473 genes (Hutchison et al., 2016). Another larger international team lead by Jeff Boeke is extending genome synthesis to the eukaryotic organism – *Saccharomyces cerevisiae*. The synthetic yeast project is referred to as Sc 2.0 and it not only extends genome synthesis to eukaryotes but also includes more rational changes to facilitate further genome engineering (Dymond et al., 2011). Now, people are also planning to synthesize the genome of more advanced organisms, even human beings (Boeke et al., 2016). Though this is still controversial, it could make a big difference to synthetic biology and be beneficial to humanity with cautious design and corresponding policies.

1.2.2 Sc 2.0 project and progress

Yeasts are unicellular eukaryotic organisms that have been identified and used for many fermentation applications. The species *Saccharomyces cerevisiae* is one of the most intensively studied. *S. cerevisiae* was the first eukaryotic genome that was sequenced, and this was followed with numerous studies on gene functions and interaction for the genome annotation (Costanzo et al., 2010; Goffeau et al., 1996). The annotation and updates are stored in the Saccharomyces Genome Database (SGD). As to the genome size, the total length is more than 12 Mbp, including 16 chromosomes with sizes ranging from 0.2 Mbp to 1.5 Mbp. Based on all the knowledge of the yeast genome and the rapidly advancing techniques in synthetic biology, the community were ready to synthesize the first eukaryotic genome. Since the genome size is 12 times that of *M. mycoides*, wide collaboration is necessary. In 2011, the leading group from John Hopkins University finished partial synthesis of a chromosome, the right arm of chromosome IX, termed as synIXR and partial left arm of synVI, termed semi-synVIL (Dymond et al., 2011). Rather than simply copying nature's coding, three novel design principles were applied for the synthesized chromosomes. First, there are some recoded sequences, for example TAG/TAA stop codon swaps and PCRtags to distinguish wild-type chromosomes and synthetic ones;

second, the introns and destabilizing elements like tRNA genes and transposons were removed from the synthetic chromosomes; third, DNA recombinase recognition sites were put in between genes for further inducible evolution, which is referred to as SCRaMbLE (synthetic chromosome rearrangement and modification by loxP mediated evolution). Based on the same design principles, global collaboration was established among groups from the USA, UK, China and later Australia and Singapore to form an international consortium for the Sc 2.0 project (Figure 1.2). Six years have passed since the first report, and the synthesis of more and longer chromosomes has been accomplished (Annaluru et al., 2014; Mercy et al., 2017; Mitchell et al., 2017; Richardson et al., 2017; Shen et al., 2017; Y. Wu et al., 2017; Xie et al., 2017; W. Zhang et al., 2017). In general, the sequence modifications did not significantly change the fitness of the synthetic strains and only a few defects were detected under stressed conditions, which can be further recovered by the SCRaMbLE system(Shen et al., 2017). Other traits of the synthetic strains were evaluated and compared with the wild type strain by analysis of chromosome 3D structures, transcriptomics and proteomics profiles, cell cycle, rDNA loci effect, etc. The second design of removing all tRNA genes from synthetic chromosomes generated some problems to the synthesis of some chromosomes, but this was solved by supplementing essential tRNA genes separately. At the same time a Neochromosome composed of all the removed tRNA genes is under construction and will be combined with other chromosomes as a final goal. In the design of the tRNA Neochromosome, another recombination site rox, which is the target site of Dre recombinase, will be placed in between the tRNA genes. Finally, a special SCRaMbLE system will be applied to the special chromosome for genome evolution. The design of the tRNA Neochromosome also reflects the “modular” strategy in synthetic biology. Further background introduction on recombinases will be given in chapter 1.3.

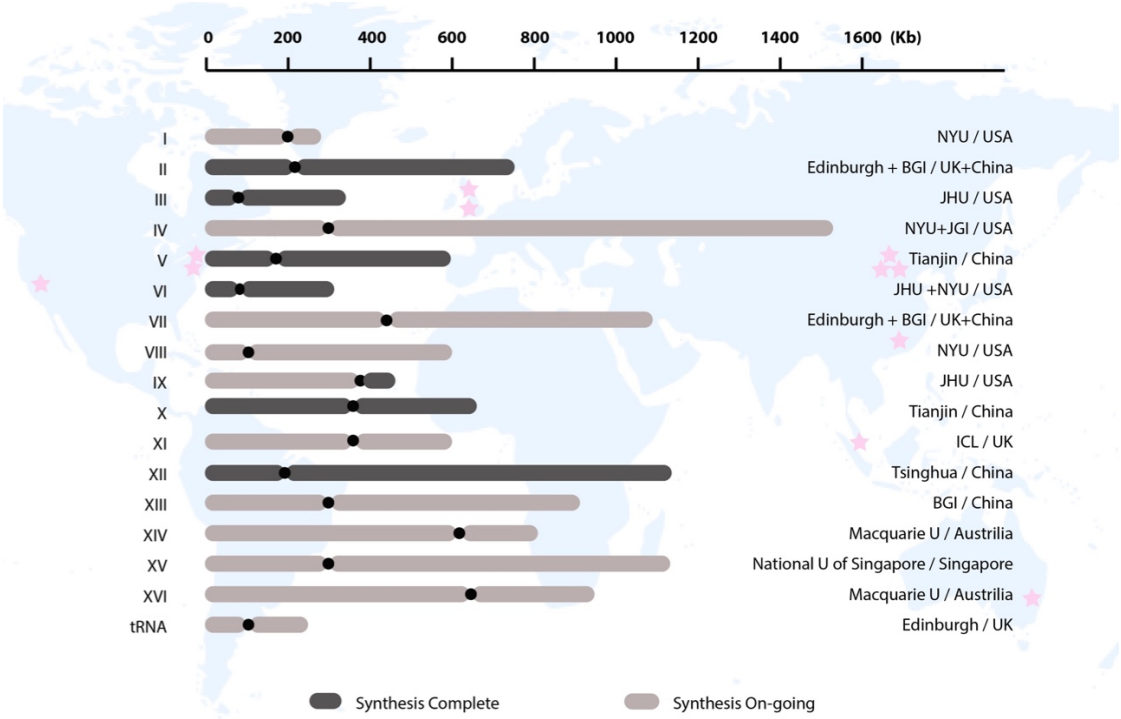


Figure 1.2| The international synthetic yeast Sc 2.0 project consortium (picture modified from the Science paper (Richardson et al., 2017))

1.2.3 Genome editing tools comparison and directed evolution

Before moving on to the subject of recombinases, it is also worth briefly reviewing other genome editing tools for comparison. There are basically two categories of genome editing tools that are specific sequence-targeted or randomly targeted. The specific sequence targeted tools include Zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and the clustered regulatory interspaced short palindromic repeat (CRISPR)-based RNA-guided DNA endonucleases (Cong et al., 2013; Joung & Sander, 2013; Urnov, Rebar, Holmes, Zhang, & Gregory, 2010). These DNA editing tools use different mechanisms to bind nucleases to specific DNA sites and generate DNA cleavage to achieve DNA disruption, addition or correction for different purposes (Gaj et al., 2013). The invention of these genome editing tools opens the door to precisely study specific gene functions in different organisms and even in specific tissues in advanced creatures. However, biological systems are huge chemical networks with complex interactions. Sometimes, single gene study is not sufficient to describe a biological function systematically. Also, the accuracy of

genome sequence recognition and cutting is still not stringent enough to avoid all off-target effects, which makes it even harder to correctly modify multiple genes simultaneously and efficiently. Another type of genome editing tool does not target specific genes but is used to generate sequence diversity randomly across the whole genome with cell replication, and this is called multiplex automated genome engineering (MAGE) (H. H. Wang et al., 2009). It accelerates the process of creating variations in new generations and with proper selection methods, strains with favorable traits under selection can be screened out, which is also called directed evolution. By MAGE, the deoxy-D-xylulose-5-phosphate (DXP) biosynthesis pathway was optimized in *E. coli* and this successfully improved lycopene production within several days (H. H. Wang et al., 2009). However, each time MAGE is applied for a different purpose, the ss-oligonucleotides need to be re-designed and re-synthesized. Furthermore, the variations were generated only in certain genes in the DXP pathway and MAGE is not yet able to generate variations across the whole genome and achieve chassis optimization. The SCRaMbLE system designed in the Sc 2.0 project includes the application of site-specific recombinase Cre/LoxP to achieve precise DNA recombination between target sites; at the same time, since there are hundreds of such sites distributed across the whole genome, it can generate theoretically infinite recombination possibilities of random variations needed for directed evolution. Compared with ZFNs, TALENs and CRISPR, the SCRaMbLE system can target multiple genes at the same time with precision; compared with MAGE, it can generate random variations across the whole genome to achieve chassis optimization. Though it needs whole genome synthesis to introduce the recombination sites, once the synthesis is accomplished, it can be a universal platform for many applications, especially in metabolic engineering and for industrial use.

Table 1.1 Comparison of genome-editing tools

	ZFNs	TALENs	CRISPR	MAGE	SCRaMbLE
DNA interaction	Protein	Protein	RNA +protein	DNA oligos	Protein
Recognition	3-6 nucleotide triplets each ZNF	Single nucleotide each domain	RNA guided	DNA oligo guided	Recombination sites
Editing type	Nuclease	Nuclease	Nuclease and single base pair switching	Base pair mismatches insertions deletions	Large fragment deletions inversions duplications
Usage	Specific sequence editing			Directed evolution of chassis	

1.3 Site-specific recombinases and integrases

Site-specific recombinases (SSRs) are a family of DNA recombination enzymes that can recognize a short target site sequence and drive DNA recombination between the sites to achieve either excision, integration or inversion of DNA fragment between the sites. The SSRs have been discovered in many organisms, mainly in bacteriophage such as Cre/loxP, Dre/rox and φC31 integrase, in yeast such as Flp/FRT, and in bacteria such as Vika/Vox, VCre/VloxP and SCre/SloxP systems (F. Guo, Gopaul, & van Duyne, 1997; Hillman & Calos, 2012; Karimova et al., 2013; Lloyd & Davis, 1994; Sauer & McDermott, 2004; Suzuki & Nakayama, 2011). The main natural function of the recombinase in bacteriophage is to stabilize and duplicate by integrating the viral genome into the host genome. The basic process for site-specific recombination includes sequence recognition, protein-DNA binding and catalysis, and DNA strand exchange. Based on the conserved nucleophilic amino acid residue they use for attacking DNA to generate a strand break, recombinases are categorized into two families: the tyrosine recombinase family and the serine recombinase family. Besides amino acid homology, the detailed mechanism and process of DNA recombination also differs for the two families, and they are introduced separately below.

1.3.1 Tyrosine recombinase

1.3.1.1 Function and mechanism

Structure of recombination site. Normally, a recombination site consists of two palindromic arms and one spacer region between the two arms. The palindromic arm is responsible for binding the recombinase and the spacer region is for homology recognition and strand exchange.

Process of recombination. At the start of the recombination, four recombinase subunits are recruited to the two recombination sites and approach each other to be close in space; after recombinase-site binding, recombinase becomes covalently linked to the DNA through the conserved catalytic tyrosine residue and forms the synaptic complex; recombination happens by exchanging one set of strands to form the Holliday junction intermediates and then finishes by exchanging the other set (F. Guo et al., 1997) (Figure 1.3a). The recombination process is all accomplished by the interaction of recombinase and corresponding site; no ATP or accessory proteins are needed for the process.

DNA behavior with recombination. The palindromic arms are symmetrical while the spacer region is normally not and the spacer sequence determines the directionality of a recombination site. When two recombination sites are located on one DNA fragment linearly, if they are in the same orientation, recombination will result in the excision of the fragment between them; if they are in the opposite orientation, recombination will result in inversion of the intervening fragment (Figure 1.3b). One special feature of this type of recombination is that the reaction is bidirectional: excision and integration are reversible and inversion is reversible.

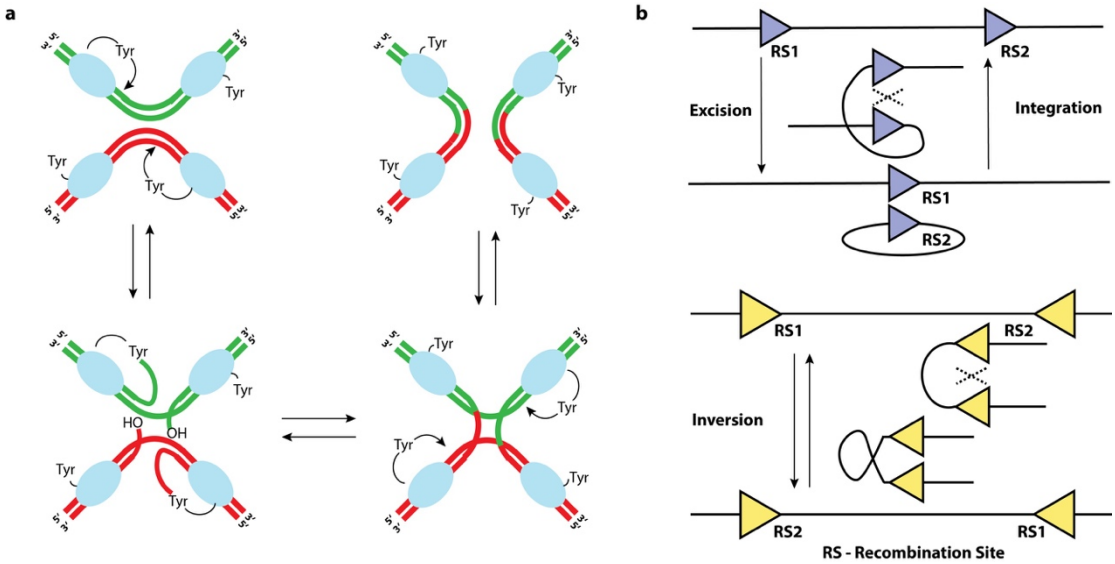


Figure 1.3|Function and mechanism of tyrosine recombinase. **a.** Recombination mechanism of tyrosine recombinases. **b.** Recombination site orientation determines the result of DNA recombination.

1.3.1.2 *In vitro* and *in vivo* function

Flp and Cre are the earliest discovered and most studied DNA tools for genome editing (Sauer, 1987). Though found in yeast and bacteriophage respectively, these recombinases have been expressed and used in other species, including mammalian cells, for genome tailoring (Nagy, 2000; Nern, Pfeiffer, Svoboda, & Rubin, 2011; Schwenk, Kuhn, Angrand, Rajewsky, & Stewart, 1998). Beside single genes, they have also been used for rapid and large scale segment swapping or insertion into the genome (Krishnakumar et al., 2014; J. G. Thomson, E. B. Rucker, & J. A. Piedrahita, 2003b). In recent years, more applications have utilized multiple orthogonal SSRs for sequential binary control of gene expression or for the design of biosafety devices to control the spread of synthetic biological tools (Y. Cai et al., 2015; Hermann et al., 2014). The crystal structures of Cre and Flp as well the structures of the SSR-DNA synapse complexes were also well studied before all these applications (Y. Chen, Narendra, Iype, Cox, & Rice, 2000; F. Guo et al., 1997), following successful extraction and purification of SSRs and subsequent *in vitro* functional studies. Taking Cre as an example, the *in vitro* function of Cre was demonstrated in the 1980s, and quantitative biochemical characterization was also reported (Ghosh & Van Duyne, 2002). The

purification method was further developed and a commercial product is even available for academic use (Cantor & Chong, 2001). The *in vitro* function was also used for studying structure and function of the recombination sites and for screening functional site mutants that will be introduced in the following section (Missirlis, Smailus, & Holt, 2006; Sheren, Langer, & Leinwand, 2007).

1.3.1.3 Engineering of the recombination system

After the natural function and mechanism of SSRs were discovered and explained, people started to explore and extend the potential of the original systems by engineering the recombinase and recombination sites.

For the SSR engineering, directed evolution methods were used to generate new recombinase variants for binding new sites (Buchholz & Stewart, 2001). Another reason for the engineering of the recombinase is that sometimes the heterologous expression of the recombinase results in toxicity to the host strain, in most cases because of the existence of pseudo-sites on host genome (Janbandhu, Moik, & Fassler, 2014; Loonstra et al., 2001). To reduce off-target effects, recombinase Cre has been engineered to improve recombination accuracy by destabilizing binding cooperativity (Eroshenko & Church, 2013).

While SSR engineering provides some new recombinase versions, there are more studies focusing on the engineering of recombination sites. Since the structure of a recombination site is categorized as spacer region and palindromic arms, the site mutants were generated and studied separately for the two regions for different purposes (Kimi Araki & Yamamura, 2012). Taking the Cre/loxP system as an example, the 34bp loxP site is composed of a 13-8-13 arm-spacer-arm structure (Figure 1.4a). The sequences of the two 13bp arms are inverted repeats while the natural 8bp spacer region is asymmetric.

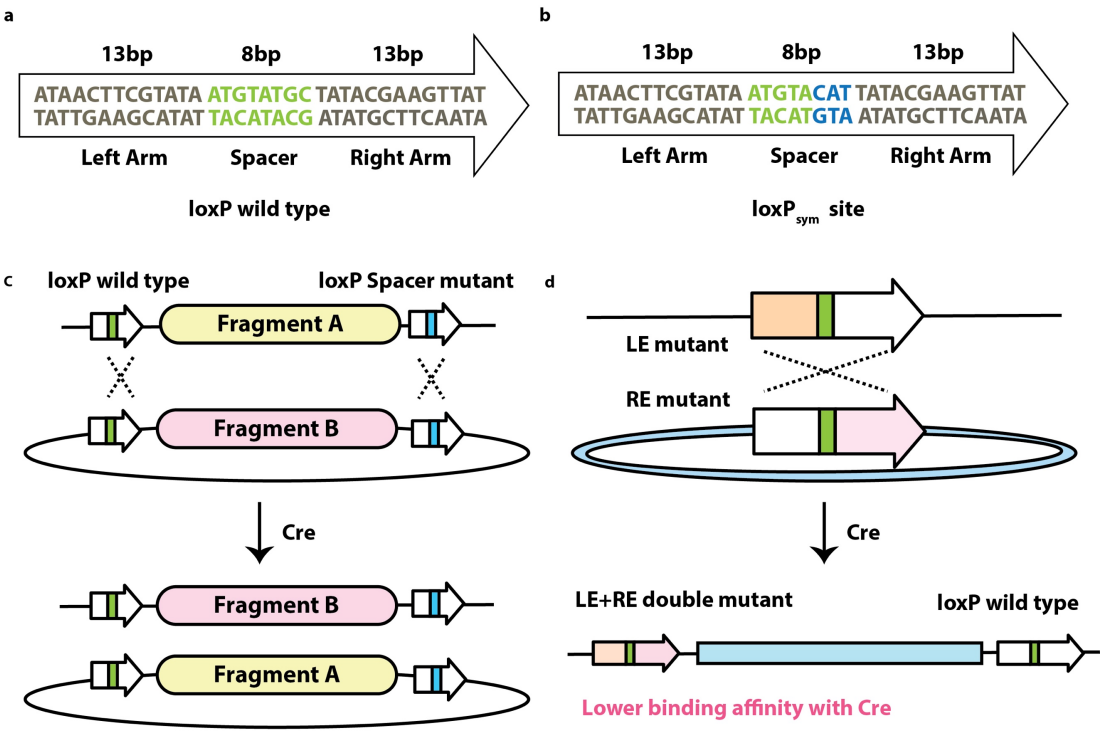


Figure 1.4| Structure of recombination site, site engineering and application. a. Structure of loxP wild type site. **b.** Structure of engineered symmetrical $loxP_{sym}$ site. **c.** Diagrammatic sketch of recombinase-mediated cassette exchange (RMCE). **d.** Illustration of LE-RE double mutant strategy.

Design of RMCE strategy. In a fundamental study on the function of the spacer region, mutant loxP sites with single base substitution mutation were created to study the homology and directionality features of the sites (Hoess, Wierzbicki, & Abremski, 1986). With single-base mutations, some loxP mutant sites orthogonal to the wild type loxP were created. Recombination can still happen within the mutant site pairs but not with loxP wild-type. After the success of 1 bp mutant sites, a 3 bp mutation site was also created with a symmetrical spacer region – the loxPsym site (Figure 1.4b). The completely symmetrical spacer enabled that DNA excision and inversion can happen in equal chance between loxPsym sites. Twenty years later, this loxPsym site was chosen as a basic element for the design of the yeast synthetic chromosomes to enable the SCRaMbLE process in the synthetic genome. Many other non-compatible loxP sites were generated later for orthogonal use in mammalian cells *in vivo* or by high-throughput screening *in vitro* (S. J. Langer, Ghafoori, Byrd, &

Leinwand, 2002; Missirlis et al., 2006). These heterospecific loxP sites were used in pairs to drive fragment exchange with target sequences in the chromosome, which is termed recombinase-mediated cassette exchange (RMCE) (Figure 1.4c) (K. Araki, Araki, & Yamamura, 2002). RMCE improved the ability of recombinase on genome engineering by facilitating the knockout and switch of target fragments in one step.

Design of LE+RE strategy. Beside excision and inversion, integration is also of great usability for inserting DNA fragments into target sites in the genome. Since excision and integration are reversible, theoretically the reaction should have equal possibility in both directions. However, the possibility of molecular collision of two recombination sites for excision and integration are different: the possibility of recombination of two sites that are not associated in the cell for integration is lower than that of two sites which are located on the same piece of DNA for excision. To solve the problem of lower integration efficiency, the arm regions of the loxP site were engineered to reduce the reversibility after integration. A left element/right element (LE/RE) strategy was applied to stabilize the insertion product (Albert, Dale, Lee, & Ow, 1995). The LE and RE refer to recombination site variants with mutations either in the left or right palindromic arm. The recombination between LE site pairs or RE site pairs is not much affected, but when the mutations in LE and RE were joint in a single "LE+RE" hybrid site, the binding affinity with the recombinase is reduced. A "LE+RE" hybrid site and wild type loxP site can be generated by the process of integration between LE and RE site (Figure 1.4d). Since the binding affinity of the hybrid site is reduced, the reverse recombination between the wild-type and hybrid site becomes more difficult, which interrupts the reversibility and promotes the reaction moving in the direction of integration. The LE+RE strategy was also combined with RMCE for stable gene knock-in in ES cells (K. Araki et al., 2002).

1.3.1.4 Regulation of recombinase function *in vivo*

One main feature of an artificial system is its controllability. It is important that the function of the recombinase is well controlled to drive the recombination process. The recombination process can be turned on by certain inputs and tuned by the expression level of the recombinase.

Transcriptional regulation. Taking *S. cerevisiae* as an example, the expression of recombinase can be tuned transcriptionally using inducible promoters like P_{GAL1}, P_{MET25} and P_{CUP1} or non-inducible promoters like P_{CYC1}, P_{ADH1}, P_{TEF2} (Cheng, Chang, Joy, Yablok, & Gartenberg, 2000), etc. The strong promoters can ensure a high expression level of recombinase and increase the recombination rate, which is especially useful for gene knock-out. However, sometimes expression driven by strong inducible promoters like P_{GAL1} has a leaky effect (Gohil, Thompson, & Greenberg, 2005), making it less tightly controlled and undesirable for computing applications. To lowering the expression, other weaker promoters like the cell-cycle related promoter P_{SCW11}, which can be activated only once in a cell cycle, can be used to control the expression level (Lindstrom & Gottschling, 2009).

Post-translational regulation. Beside transcriptional regulation, a ligand-based post-translational regulatory strategy can also be applied to achieve tighter control of recombinase function. In natural cell systems, there are a class of nuclear receptors that can sense hormones and certain other molecules to regulate the expression of specific genes (Evans, 1988). This type of nuclear receptors is also a type of transcription factors, with DNA binding ability and transcription-associated functions. Based on these two features, the functional domains of a nuclear receptor can be divided to a DNA-binding domain (DBD) and a ligand-binding domain (LBD) connected by other terminal domains and hinge domains. The binding of a small molecule with the nuclear receptor can cause a conformational change in the receptor and trigger subsequent up- or down-regulation of gene expression (Kumar & Thompson, 1999). The nuclear receptors are mainly subdivided into four types. Those of type I are the best studied and have been applied to recombinase regulation. In the absence of ligand, the nuclear receptor is bound by heat shock proteins (HSPs) in the cytosol; with the binding of a certain kind of ligand, the nuclear receptor is released from the HSPs and translocated from the cytoplasm to the nucleus for transcription function(Linja et al., 2004). Since the recombinase is a DNA binding protein, it is feasible to simulate the natural design and fuse the recombinase with the LBD to achieve spatial control. This idea was first successfully achieved with FLP fused with

estrogen-binding domain (EBD) in a mammalian cell line (Logie & Stewart, 1995). Later, more successful pairs of fusion proteins were reported, including Cre and progesterone-binding domain (PBD) or Dre and PBD regulated by synthetic steroid RU486 in mammalian cells (Anastassiadis et al., 2009; Kellendonk et al., 1996; Wunderlich, Wildner, Rajewsky, & Edenhofer, 2001). The fusion protein strategy not only works in mammalian cells where the natural system exists, but also works well in yeast with one example being the fusion of Cre-EBD applied to a mother cell enrichment program in *S. cerevisiae* (Lindstrom & Gottschling, 2009). In addition to fusing with recombinase, the LBDs have also been engineered for the design of artificial transcription factors (ATFs) to tune gene expression like that of the recombinase transcriptionally (McIsaac et al., 2013; Quintero, Maya, Arevalo-Rodriguez, Cebolla, & Chavez, 2007). In addition to protein fusion with the LBD of a nuclear receptor, recombinases like Cre have also been split into two moieties to fuse with proteins in ligand-induced dimerization systems for function regulation (Jullien, Sampieri, Enjalbert, & Herman, 2003).

1.3.2 Serine recombinases

1.3.2.1 Function and mechanism

Though tyrosine recombinases are the main focus of my PhD project, the serine recombinases are also worth mentioning for comparison and possible future applications. Serine recombinases are also site-specific DNA recombinases but are different from the tyrosine recombinase in two ways: the length and composition of recombination site pairs and it is unidirectional without accessory protein. Among the serine recombinases, the subgroup of serine integrases is of most interest. There are four sites involved in the process of recombination, named originally from the attachment process of a phage site to a bacterial host site: *attP* and *attB.* The integration of *attP* to *attB* will generate two new sites called *attL* and *attR*. The integration reaction of this type of recombination is irreversible unless a recombination directionality factor (RDF) is introduced (Smith, Brown, McEwan, & Rowley, 2010) (Figure 1.5a). Only with the presence of an RDF can fragment excision

happen between the *attL* and *attR* sites, which makes it ideal for unidirectional integration and controllable excision. The molecular behavior has also been explained structurally by modeling the interaction between serine recombinase, DNA site and RDF (Rutherford, Yuan, Perry, Sharp, & Van Duyne, 2013). Likewise, the *in vitro* function of serine recombinases has also been demonstrated and used to study the structure and function of the recombination site by testing recombination efficiency of target site mutants (McEwan, Rowley, & Smith, 2009; Singh, Ghosh, & Hatfull, 2013). An *in vitro* mechanistic model was also developed to describe the *in vitro* recombination efficiency and predict future experimental data (Bowyer, Jia, Rosser, Colloms, & Bates, 2015).

1.3.2.2 Engineering and applications

The function of many serine recombinases has been tested in *E. coli*, mammalian cells and *in vitro*, with representatives like Bxb1, ϕC31, R4, ϕFC1, and ϕRV1 integrases, etc. (Keravala et al., 2006; Thorpe & Smith, 1998). These serine recombinases have been applied to build genome integration tools for mammalian synthetic biology, and Bxb1 integrase was reported to be the best candidate for mammalian genome engineering (Duportet et al., 2014; Xu et al., 2013). The Bxb1 and ϕC31 integrases were also utilized to build logic-gate based cell computing and memory devices (Siuti et al., 2013). Because of the irreversible recombination process, recombinase-based logic gates can be more tightly controlled compared with traditional inducible promoter-based logic gates which are probably associated with leaky effects. Moreover, the effect of the input can be stably memorized and stored by the cell and passed on to the next generations, which is a milestone in the development of logic gates. In addition to *in vivo* applications, the serine recombinases have also been used *in vitro* for developing a new method for the assembly of metabolic pathway circuit (Colloms et al., 2014b). To enable multiple-part assembly, orthogonal recombination sites of *attP* and *attB* were engineered at the central 2 bp which needs to be perfectly matched to enable recombination (Figure 1.5b). Due to time limitation and the compatibility with the design of Sc 2.0 synthetic chromosome, the application of

serine recombinases was not included in this project. However, the serine recombinases could be a good substitute for *in vitro* use of the tyrosine recombinases and a good substitute for *in vivo* use to design Neochromosomes with the serine recombinase recognition sites in future.

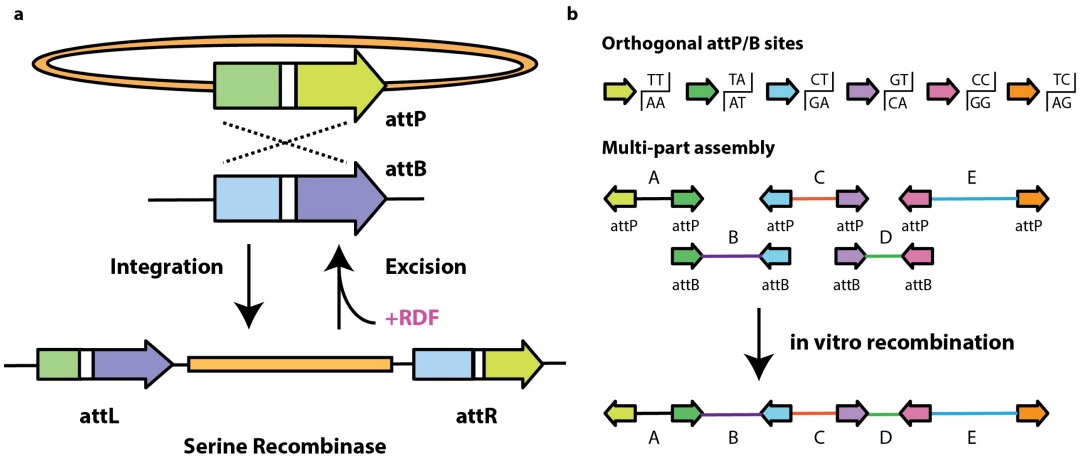


Figure 1.5|Function and application of serine integrases. a. Recombination mechanism of serine integrases. **b.** Multi-part DNA assembly by applying orthogonal serine recombination sites.

1.4 Natural products and synthetic biology

1.4.1 Overview of synthetic biology and metabolic engineering

Another fast growing aspect of synthetic biology is its coupling with metabolic engineering for use in industrial biotechnology to produce biofuels, medicine and high value chemicals like nutrition supplements, anti-tumor molecules, and antibiotics (Keasling, 2012; Stephanopoulos, 2012; Yadav, De Mey, Lim, Ajikumar, & Stephanopoulos, 2012). These special metabolites with biological activities are very important for clinical uses and therefore a lot of effort has been made to achieve high production at industrial scale or to discover new drugs for treating various diseases. The natural resource of chemicals is enormous, and the natural products discovered and utilized are just the tip of the iceberg. However, many chemicals are produced at very low levels, making them difficult to detect. Even though many chemicals are screened out from nature, the species producing them is sometimes not cultivable under laboratory conditions, let alone in industrial fermentation. In traditional

metabolic engineering, metabolic analysis, substrate optimization and direct genetic modification were the most common methods for improving the yield of target chemicals (Nielsen & Keasling, 2011). In many cases, the traditional methods of engineering natural pathways by small scale modification are still limited. Because of the complex cellular networks and the lack of knowledge of specific gene activation, the production process of target chemicals is rather uncontrollable.

The advent of synthetic biology provides the opportunity to engineer the pathways more precisely and systematically. With the continuously decreasing cost of DNA synthesis and invention of diverse assembly methods, it is feasible and becoming easier to design and construct pathways and transfer them to specific hosts for heterologous expression (DeLoache et al., 2015; Hawkins & Smolke, 2008; Kosuri & Church, 2014; Paddon & Keasling, 2014b; Thodey et al., 2014). The engineering process of a metabolic pathway was recently reviewed in 2016, including pathway design, construction and screening (Smanski et al., 2016). The first and core part in pathway design is to determine what genes are to be used for natural product synthesis. With the increasing volume of genome sequencing data, researchers found that genes encoding the synthesis of a natural product are usually colocalized in compact biosynthetic gene clusters (BGCs). Through genome mining and homology BLAST, many potential BGCs from diverse species or mixed communities like the various microbiota are accessible (Donia & Fischbach, 2015). With such large DNA resources and well characterized genetic regulatory circuits, it is possible to construct large DNA libraries for screening natural products with desired function. Large-scale construction technologies are being developed to continuously achieve automated high-throughput DNA fabrication for parallel combinatorial biosynthesis, making it more efficient and less error prone than manual construction. On top of this, more elaborate engineering methods have been applied for tuning specific pathways.

1.4.2 Tuning methods for heterologous metabolic pathways

Through the first step of genome mining, targeted BGCs for pathways can be refactored for heterologous expression. For successful heterologous expression, it is necessary to fine-tune the expression and function of each gene in a new cellular

system. It is not the case that every gene should be expressed as strongly as possible to maximize production. From the perspective of transcriptional regulation, many regulatory devices have been discovered and engineered in bacteria and yeast to obtain enough regulatory parts with a wide range of transcription initiation strengths and some special functions like ligand-inducible promoters (Liang, Ning, & Zhao, 2013), synthetic promoters (Redden & Alper, 2015), stress-response promoters (Dahl et al., 2013), etc. Various combinatorial assembly tools have been used to shuffle the engineered regulatory parts with different BGCs for screening for pathways with the best balanced expression (Du, Yuan, Si, Lian, & Zhao, 2012; Jones et al., 2015; Lee, Aswani, Han, Tomlin, & Dueber, 2013; Mitchell et al., 2015; Smanski et al., 2014). But sometimes there are too many genes in a pathway and the number of possible combinations is too high to cover. On one hand, a process called multivariate modular metabolic engineering that groups subsets of genes together has been applied for combinatorial assemblies with gene groups (Biggs, De Paepe, Santos, De Mey, & Kumaran Ajikumar, 2014). While it works in bacteria, it is less suited to use it in eukaryotes in which all genes are monocistrons. On the other hand, feedback and feedforward loops were designed to dynamically respond to the production of target chemicals or intermediates by triggering of an efflux pump (Tahlan et al., 2007), chemically inducible/repressible promoters or RNA aptamer sensors (Michener, Thodey, Liang, & Smolke, 2012; Siddiqui et al., 2012). With the help of fluorescent reporters, cell sorting can be used to select the clones with higher production from the library more efficiently. Besides engineering of gene regulation, protein engineering is sometimes also necessary since folding and post-translational modification might be affected in the engineered organisms. Random mutagenesis or targeted mutagenesis have been used to generate protein variants for screening (DeLoache et al., 2015; Marcheschi, Gronenberg, & Liao, 2013). However, this can sometimes be difficult since mutagenesis will not always guarantee successful screening and it relies heavily on whether a structural model is available or not for the right mutation targets. Therefore, it is still difficult to rationally design or change certain domains in the target proteins. Currently, alternative ways to deal with the

problem can be switching and screening for analogs of critical enzymes from other organisms. For example, Smolke's group successfully completed the biosynthesis of opioids in yeast by incorporating enzyme genes from bacteria, yeast, plant and even mammals (Galanie et al., 2015). Besides enzyme engineering, the use of a protein scaffold was also applied in the case of artemisinin biosynthesis to tether functionally related enzymes in an assembly line to improve the production (Dueber et al., 2009). The engineering of the composition of the pathway itself is straightforward, while the host context can also make a difference since the heterologous pathway can also interact with cellular networks and add burden to the host cell by competing for cell resources and energy or even generating intermediates toxic to the host. Different methods for measuring and reducing cell burden were reviewed (G. Wu et al., 2016), among which chassis optimization is one of the most important aspects. Genome-scale engineering method including MAGE (H. H. Wang et al., 2009) (multiplex automated genome engineering) and TRMR (Warner, Reeder, Karimpour-Fard, Woodruff, & Gill, 2010) (trackable multiplex recombineering) were used to either improve the production of lycopene quickly or generate faster growing strains under industrially related conditions in *E. coli*.

In addition to the above methods, other directions like engineered symbiotic microbial consortia (Zhou, Qiao, Edgar, & Stephanopoulos, 2015), and cell-free metabolic engineering (Dudley et al., 2015) are also promising platforms for synthetic natural product applications.

1.4.3 Benchmark pathways – β-carotene and violacein pathways

Among the huge amount of natural product producing pathways, the pigment generating pathways are good benchmark pathways for different research purposes with their advantage of visibility. In this research, two pigment generating pathways, the β-carotene and violacein pathways were used for demonstrating pathway diversification in different chapters. β-carotene is a red-orange pigment that is abundant in plants and it belongs to the large and diverse chemical class of terpenoids. β-carotene was viewed as a good food supplement that can be further catalyzed to vitamin A and therefore it attracted the attention from both industries

and researchers. Since the production of carotenoids is limited by the low yields and difficult extraction from plants, microbial production and fermentation has been investigated and attempted. It has been reported that carotenoids can be synthesized in yeast strains and various productions were observed in these different strains (Mata-Gomez, Montanez, Mendez-Zavala, & Aguilar, 2014). In yeast strains like *S. cerevisiae* and *X. dendrorhous*, the precursor of carotenoids, farnesyl diphosphate (FPP), can be synthesized from acetyl CoA by local yeast enzymes; then three heterogenous genes, *CrtE*, *CrtYB* and *CrtI*, were introduced for following conversions towards β-carotene (Verwaal et al., 2007). FPP is converted to geranylgeranyl diphosphate (GGPP) by the GGPP synthase encoded by *CrtE*; two GGPP molecules can be catalyzed to phytoene by the bifunctional enzyme CrtYB; the colorless phytoene is converted to the yellow color neurosporene first and then to the red color lycopene through four desaturation reactions by the CrtI desaturase; finally, enzyme CrtYB plays the other role as cyclase to convert lycopene to the orange color β-carotene (Figure 1.6a). Besides β-carotene pathway, the purple antibiotic generating pathway – violacein pathway – also attracted commercial and academic interests. Violacein is a bisindole pigment originally produced in various genera of bacterial strains, including *Collimonas*, *Janthinobacterium*, *Microbulbifer sp.*, etc (Choi, Yoon, Lee, & Mitchell, 2015). The key genes essential for violacein synthesis have been identified as *VioA, B, C, D, E* gene cluster in *Chromobacterium violaceum* by DNA sequencing, mutagenesis and chemical analysis (August et al., 2000). The five genes have also been sub-cloned and proven to be functional to synthesize violacein in other bacterial strains like *E. coli* and eukaryotic microorganism like *S. cerevisiae* (Jones et al., 2015; Rodrigues et al., 2013). The violacein pathway starts with the transformation of two substrate molecules of L-tryptophan into indole-3-pyruvic acid imine (IPA imine) by VioA; through two steps catalysis by VioB and VioE, IPA imine is first turned into IPA imine dimer and then to the protodeoxyviolaceinic acid; the protodeoxyviolaceinic acid is further oxidized to protoxyviolaceinic acid by VioD with the presence of NADPH, H^+ and O_2; the protoxyviolaceinic acid was further oxidized to the violaceinic acid by VioC with

NADPH, H^+ and O_2; after the above five enzymatic steps, the violaceinic acid is converted to the purple violacein with a final non-enzymatic step by auto-oxydization (Lee et al., 2013) (Figure 1.6b). Besides the main flux towards violacein, some side intermediates and side products are also generated through non-enzymatic oxidization or the function of VioC, which includes the pink deoxyviolacein, the olive green prodeoxyviolacein, the pine green proviolacein and some other derivate, the reaction mechanism of which is still not clear. Therefore, the production of violacein in natural host is normally not pure and the color of host indicates the proportion of violacein to other intermediates or side products. Considering their relative simplicity, visibility and availability, the two benchmark pathways were chosen to demonstrate pathway diversification capability of the recombinase toolkits.

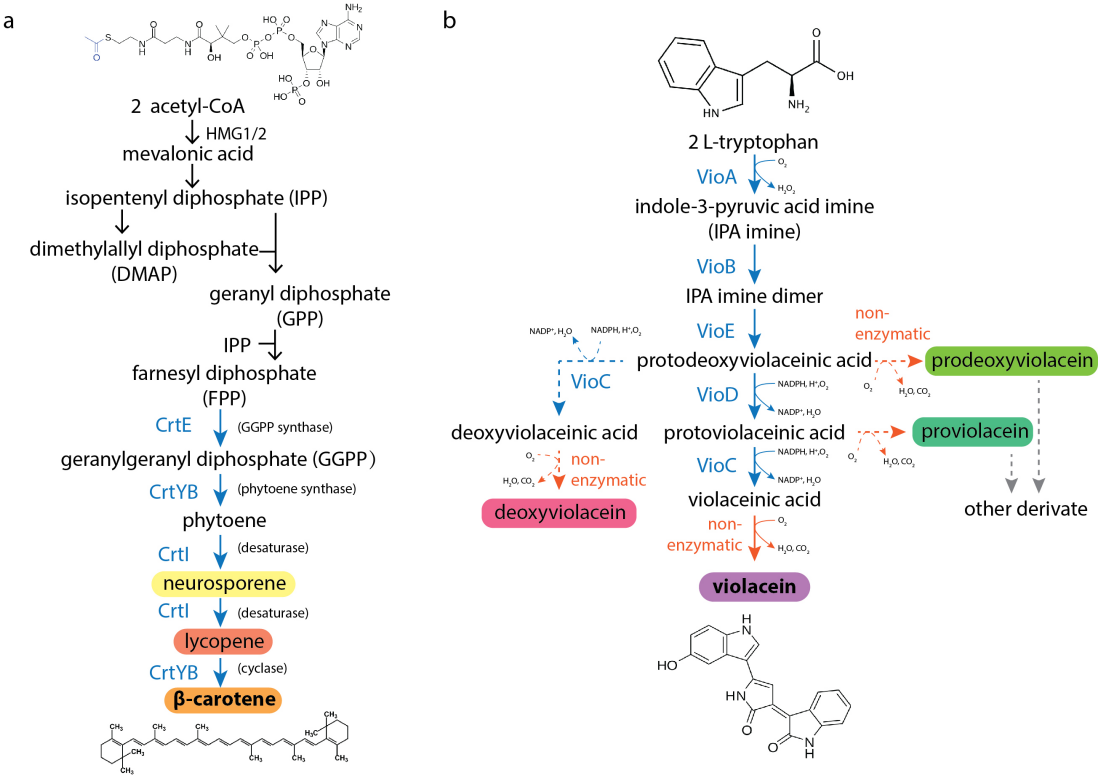


Figure 1.6| Overview of the β-carotene and violacein pathways. a. The pathway map of β-carotene synthesis. **b.** The pathway map of violacein synthesis.

1.4.4 Synthetic natural products for drug discovery

One major application field of natural products is medical uses for human health. Antibiotics are one of the vital drug discoveries benefiting human lives by curing various infections. However, the overuse of antibiotics from inappropriate

prescribing and extensive agricultural use has resulted in the crisis of antibiotic resistance (Ventola, 2015). At the same time, the number of new drugs available is dropping constantly in recent years, making the situation even worse. Much research effort has been applied to studying the mechanisms of antibiotic resistance and developing new medicines. Antibiotic resistance is normally due to genetic mutations that can lead to quite diverse mechanisms to escape the function of antibiotics. The evolutionary rate of microorganisms is much faster than the speed of new drug discovery. Screening from natural resources and chemical synthesis are the main methods for new drug discovery. However, the screening process is restricted by low level production and fermentation problems as mentioned above, and chemical synthesis is of high cost and hard to achieve sometimes because of the lack of proper catalysts. Synthetic biology provides an engineering platform that can generate a huge amount of genetic diversity which can potentially result in chemical diversity. This will accelerate the speed of genetic evolution of natural product-producing pathways and provide large pools of chemical candidates for treating different diseases.

Among infectious diseases, tuberculosis (TB) treatment is one of the most challenging problems the world is facing. TB is caused by the bacterial pathogen *Mycobacterium tuberculosis* and the treatment of TB requires drug treatment for more than 9 months, which is not affordable for many patients in developing countries (Zumla, Nahid, & Cole, 2013). The poor availability of drugs not only leads to death but also antibiotic resistance. Multidrug-resistant TB (MDR-TB) and extensively drug-resistant TB (XDR-TB) species have emerged and are spreading. Therefore, it is urgent to discover complementary new drugs that can treat such strains to save lives and constrain the spread. To establish a synthetic biology platform for anti-TB drug discovery, Professor Mike Tyers' group in Canada initiated a project that utilizes hundreds of biosynthetic enzyme genes (BSGs) from various microorganisms combinatorically to synthesize and diversify natural product (NP)-like compounds against TB, in collaboration with our group and another group in Canada to form a three-partner consortium. Based on previous research, five NP classes were chosen for BSG library construction,

including non-ribosomal peptide synthases (NRPSs) based class, polyketide synthetases (PKSs) based class, ribosomally synthesized and post-translationally modified peptides (RiPPs) derived class, nucleosides class and flavonoids class. There are many molecules from these classes that have reported antibiotic activity including viomycin, penicillin, vancomycin and kirromycin from NRPSs class (Nolan & Walsh, 2009), rapamycin and tyrocidine from PKSs class (Mootz & Marahiel, 1997; Park, Yoo, Ban, & Yoon, 2010), Nocathiacin from RiPPs class (Ding et al., 2010; Pucci et al., 2004), capuramycin from nucleosides (W. Cai et al., 2015), etc. Since antibiotic resistance is due to the failure of the existing compounds to target and interrupt normal physiological processes of a certain pathogen, it is possible that structural and molecular changes of the antibiotic compound can enable recapture of the pathogen. Therefore, modifying the basic chemical scaffold by combinatorial use of various tailoring enzymes is a very promising way to generate enormous chemical diversity.

Yeast as a basic synthetic biology model organism was chosen as the heterologous pathway expression host in this project. Many flexible DNA cloning shuttle vectors are available to accommodate large synthetic fragments and as a eukaryotic organism it has more complete protein modification machinery to improve the chance of proper folding and function of target enzymes. Many drugs like artemisinin, opioids and penicillin have been successfully produced in yeast, proving the feasibility for using yeast in this synthetic natural product project (Galanie et al., 2015; Gidijala et al., 2009; Paddon & Keasling, 2014a; Thodey et al., 2014). In the three-partner consortium, different expertise like yeast molecular genetics, natural product biosynthesis and DNA synthesis and assembly automation were brought together to tackle the worldwide problem of development of new antibiotics.

1.5 Cell-free synthetic biology

Though the topic of my PhD focuses on the application of recombinase to metabolic pathway engineering, it also involves the cell-free use of recombinase and it is worth introducing the current development of cell-free synthetic biology. The idea of cell-free systems dates back to the 1890s when Eduard Buchner used yeast extract to convert sugar to ethanol and carbon dioxide. Later on, with deeper understanding

about cellular systems in modern biological studies, people were able to purify single components for *in vitro* use, like the various DNA polymerases and restriction enzymes. Now, researchers are aiming to put together more elements *in vitro* to achieve more complex networks outside the cell. Combined with synthetic biology, gene circuits and networks can be potentially expressed and function outside the cell. The methods to create a cell-free system are divided into crude extract cell-free systems (CECFs) and synthetic enzymatic pathways (SEPs) (Hodgman & Jewett, 2012). Many applications of cell-free systems have been reviewed, including protein and peptide synthesis, metabolite synthesis, programming circuits like genetic oscillators (Niederholtmeyer et al., 2015), compartmentalization, spatial organization, minimal cells and brand new chemical design and assembly (Hodgman & Jewett, 2012), etc. In recent years, chip-based or paper-based cell-free biosensors have also opened a new direction for developing cheap and rapid medical diagnostic tools (Cate, Adkins, Mettakoonpitak, & Henry, 2015; Pardee et al., 2016; Slomovic et al., 2015; Yetisen, Akram, & Lowe, 2013).

In addition to *in vivo* applications of synthetic biology for metabolic engineering, cell-free metabolic engineering (CFME) provides an alternative platform for natural product synthesis. In the case of CFME, both purified enzyme systems and crude cell lysates were used to successfully activate long enzyme pathways, including the glucose to ethanol conversion pathway (Guterl et al., 2012), non-oxidative glycolysis pathway (Bogorad, Lin, & Liao, 2013), hydrogen generating pathways (Martin del Campo et al., 2013), isoprenoids biosynthesis pathways (X. Chen et al., 2013; Harper, Bailey, Edwards, Detelich, & Keatinge-Clay, 2012; Zhu et al., 2014) and high-value therapeutic pathways (Hagen et al., 2016; Harper et al., 2012; Kim, Kang, Chae, Park, & Choi, 2000). The advantages of CFME include: it can be used for debugging and optimizing certain biosynthetic pathways by precise functional characterization of the catalytic enzymes; it can break the limits caused by cellular toxicity of products or intermediates and competition with other local pathways resulting from restricted energy resources; it is also faster since it avoids cell culture and accelerates the speed of design-build-test cycling (Dudley et al., 2015). Though promising, there are still

challenges in the field like the high cost of protein purification, stability and longevity of the purified enzymes and reactions, and immature conditions to scale up for industrial use. Therefore, *in vivo* and *in vitro* work should be developed at the same time and be applied complementarily to specific projects.

1.6 Thesis statement

1.6.1 Goal of my PhD research

My PhD project is mainly aimed towards developing recombinase toolkits for both *in vivo* and *in vitro* application in metabolic engineering based on the synthetic biology platform and methods (Figure 1.7). It incorporates both yeast genetics and *in vitro* recombination systems for metabolic pathway diversification and optimization on both pathway level and chassis level. Two pigment generating pathways, the carotenoid synthetic pathway and violacein synthetic pathway, were used for demonstration of different aspects of the research. Typical synthetic biology reporter fluorescent proteins were also used for quantification and evaluation of the gene expression diversity.

1.6.2 Thesis organization and chapter connection

In chapter 1, all the related background is introduced including the fundamental knowledge of synthetic biology including basic design principles and various DNA assembly methods, the current status of the international synthetic yeast chromosome project Sc 2.0, the structure and function of recombinases, the combination of synthetic biology with metabolic engineering, and cell-free synthetic biology. In chapter 2, the materials and methods of culture conditions, cloning methods and characterization assays for different genetic circuits are described. Chapter 3 focuses on developing different orthogonal recombinase devices *in vivo* for serving the SCRaMbLE purpose of normal synthetic chromosomes and a tRNA Neochromosome. Based on successful recombinase function *in vivo*, a SCRaMbLE-in device was designed to optimize the synthetic yeast chassis for heterologous metabolic pathways by directed evolution, which is described in chapter 4. In chapter 5, the *in vitro* recombinase toolkit is described for pathway level diversification. In chapter 6, my contributions to the collaborative SynNP project for discovery of new

antibiotics against TB is introduced. Finally, in chapter 7, the results are summarized and discussed with additional future perspectives for the projects.

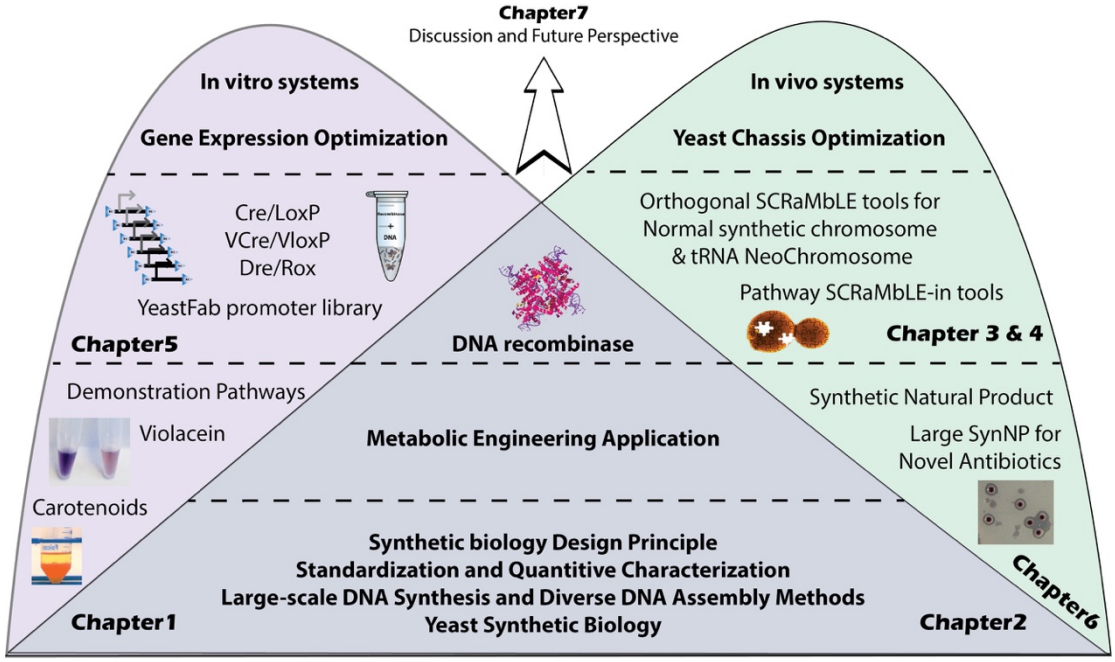


Figure 1.7 | Diagrammatic illustration of thesis structure and chapter connection.

Chapter 2 Materials and methods

2.1 Bacterial and yeast strains and culture conditions

The bacterial strains include *E. coli* strains: DH5α was purchased as competent cells from Thermo Fisher Scientific and was used for routine cloning; *ccdB* Survival™ 2 is a *ccdB* protein-resistant strain and was used for cloning DNA circuits with the *ccdB* gene; BL21 (DE3) was obtained from Jon Marles-Wright laboratory (University of Newcastle) as competent cells and was used for protein expression and purification. *E. coli* strains were cultured at 37°C except that BL21 was cultured at 18°C for protein expression. *Bacillus subtilis* and *Mycobacterium smegmatis* were obtained from the Gerry Wright laboratory (McMaster University) and used for antibiotic function assay. *B. subtilis* was cultured at 30°C and *M. smegmatis* was cultured at 37°C.

The yeast strains include BY4741 and BY4742 which are S288C derivatives with different mating types and genotypes (Brachmann et al., 1998). They were used for function testing assays for various DNA constructs. The synthetic strains were used for either co-SCRaMbLE with tRNA array, or for SCRaMbLE and SCRaMbLE-in for pathway diversification. Yeast strains were cultured at 30°C except where otherwise stated.

The genotypes of all the bacterial and yeast strains are listed in Table 2.1.

Table 2.1 Basic bacterial and yeast strains

Strain Name	**Species**	**Genotype**
DH5α	*E. coli*	*F– Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1*
ccdB Survival™ 2	*E. coli*	*F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araΔ139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG fhuA::IS2*

BL21 (DE3)	*E. coli*	*F^- ompT gal dcm lon hsdS_B($r_B^-m_B^-$) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [$malB^+$]$_{K-12}$(λ^S)*
Bacillus subtilis 168 (Ehrenberg) Cohn (ATCC® 23857™)	*Bacillus subtilis*	*ind- tyr+*
Mycobacterium smegmatis (Trevisan) Lehmann and Neumann (ATCC® 607™)	*Mycobacterium smegmatis*	NA
BY4741	*S. cerevisiae*	*MATa, his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*
BY4742	*S. cerevisiae*	*leu2Δ0 lys2Δ0 ura3Δ0 his3Δ1*
SynII	*S. cerevisiae*	*MATa leu2Δ0 met15Δ0 LYS2 ura3Δ0 his3Δ1 SYNII*
SynIII	*S. cerevisiae*	*MATa leu2Δ0 met15Δ0 LYS2 ura3Δ0 his3Δ1 SYNIII*
SynII SynXII	*S. cerevisiae*	*MATa his3Δ1 leu2Δ0 ura3Δ0 MET15 + LYS2 synOJ::HIS3*
SynII-III	*S. cerevisiae*	*MATalpha leu2Δ ura3Δ0 his3Δ1 LYS2 MET15 HO::SUP61*
SynIII VI IXR	*S. cerevisiae*	*MATalpha leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII SYN.SUP61::HO synVI SYN-WT.PRE4 IXL-synIXR*
SynV	*S. cerevisiae*	*MATa His3Δ1 leu2Δ0 met15Δ0 ura3Δ0 LYS2 synV*
SynX	*S. cerevisiae*	*MATa His3Δ1 leu2Δ0 met15Δ0 ura3Δ0 SYN10 HO::tR(ccu)J-URA3*
SynV-X	*S. cerevisiae*	*MATalpha His3Δ1 leu2Δ0 met15Δ0 ura3Δ0 LYS2 synV SYNX HO::tR(ccu)J*

2.2 Media and reagents

2.2.1 Bacterial media

LB medium was prepared by dissolving 10 g Bacto-tryptone, 5 g yeast extract and 10 g NaCl in 800 ml ddH_2O with pH adjusted to 7.5. After making up to 1 L with ddH_2O, the medium was sterilized by autoclaving and stored at room temperature. The 1000X ampicillin solutions and 1000X kanamycin solutions were filter sterilized through a 0.22 μm filter and stored in 1 ml aliquots at -20°C. The 1000X antibiotics solution were used for making LB media with 100 mg/L ampicillin or 50mg/L kanamycin.

SOC medium was prepared with 20 g Bacto-tryptone, 5 g Bacto Yeast Extract, 2 ml of 5 M NaCl, 2.5 ml of 1M KCl, 10 ml of 1 M $MgCl_2$, 10 ml of 1 M $MgSO_4$ and 20 ml of 1 M glucose. It was adjusted to 1 L with ddH_2O and sterilized by autoclaving. SOB medium was prepared with 20 g Bacto-tryptone, 5 g Bacto yeast extract, 0.58g NaCl and 0.19 g KCl. It was adjusted to 1 L with ddH_2O and sterilized by autoclave.

CCMB medium for making *E. coli* competent cells was prepared with 11.8 g $CaCl_2 \cdot 2H_2O$, 4 g $MnCl_2 \cdot 4H_2O$, 2 g $MgCl_2 \cdot 6H_2O$, 1 ml potassium acetate (5 M) and 100 ml glycerol. It was adjusted to 1 L with ddH_2O and sterilized by filtering through a 0.22 μm filter with a vacuum pump.

The *B. subtilis* overlay medium was 75 mM phosphate buffered agar overlay medium. It was normally prepared in 100 ml by mixing 0.8 g Difco Nutrient Broth, 0.40 g $NaH_2PO_4 \cdot H_2O$ (2.9 mmol), 0.65 g Na_2HPO_4 (4.6 mmol) and ddH_2O. The pH was adjusted to 7.0 and 0.7 g agar was added afterwards for autoclaving.

The Mueller Hinton Agar (MH plates) was used for the disc diffusion assay. It was purchased from Sigma-Aldrich and prepared from powder as instructed.

2.2.2 Yeast media

The YPD and YP-Galactose media were prepared with 5 g yeast extract, 10 g bacto-peptone, 10 g dextrose (glucose)/galactose, 0.16 g tryptophan, and adjusted with ddH_2O to 500 ml for autoclaving. The YPG (3% glycerol) medium was made by adding

25ml autoclaved 60% glycerol stock to 250 ml autoclaved 2X YP liquid media, then adding autoclaved 4% agar or ddH_2O to 500 ml for either plates or broth use.

The YPD medium can be supplemented with 500 μl 1000X (200 mg active/ml) G418 solution to 500ml YPD medium for antibiotic resistance gene selection. A special 2X YPAD medium was used for yeast competent cell preparation and was made by mixing 10 g Bacto yeast extract, 20 g Bacto peptone, 20g glucose and 40 mg adenine hemisulfate, and adjusted with ddH_2O to 500 ml for autoclaving.

The yeast synthetic drop-out media are commonly used for different purposes. To prepare different types of synthetic drop-out media, an amino acid synthetic complete-8 (-Ade-Cys-His-Leu-Lys-Met-Trp-Ura) dropout mix (SC-8) was first made as a master mix. The 2 X SC-8 broth was made by mixing 8.5 g yeast nitrogen base w/o added amino acids and ammonium sulfate (Sigma), 25 g ammonium sulfate (Sigma) and 10 g SC-8 dropout mix and adjusted to 2500 ml with ddH_2O. This was mixed well and distributed as 250 ml into each of ten 500 ml bottles and autoclaved for 20 min. Taking SC-His-Ura plate as an example, 250 ml 2X SC-8 broth and 250 ml 4% autoclaved agar with 25 ml 40% glucose were first mixed, and supplemented with 5 ml 40 mM Tryptophan, 10 ml 100 mM Leucine, 5 ml 15 mM Adenine, 5 ml 100 mM Lysine and 6 ml 50 mM Methionine amino acid solution (all purchased as powders from Sigma).

The counter selection 5-fluoroorotic acid (5-FOA) medium (1 mg/ml) was prepared by adding 0.5 g 5-FOA to 500 ml target media and incubated in a 50°C water bath for 30min till totally dissolved for plate or broth use.

2.2.3 Solutions and reagent kits

2.2.3.1 DNA related kits and enzymes

All DNA samples and DNA modifying enzymes were stored at -20°C and DNA purification kits were stored at room temperature. The name and vendor of the commercial kits and reagents are listed in Table 2.2. Besides the commercial chemicals and kits, other buffers and solutions were prepared in the laboratory.

The 50X TAE buffer for electrophoresis was prepared with 2 M Tris-base, 57.1 ml glacial acetic acid, 0.05 M EDTA. 242 g Tris-base was dissolved in around 600 ml ddH_2O, 57.1 ml glacial acetic acid and 0.05 M EDTA were added and mixed well. The buffer was further made up to 1 L with ddH_2O and the pH was adjusted to 8.3 for autoclaving. For a 1X working stock, 20 ml of 50X TAE was diluted in 1 L ddH_2O for use.

0.8 % TAE agarose gel was prepared with 0.8 g of agarose and 100 ml 1X TAE buffer in a 250 ml conical flask. The agarose was dissolved by heating in a microwave for 2.5 min and poured into the Bio-Rad molds with proper combs.

The 5X Isothermal buffer (6ml) for Gibson assembly master mix was prepared with 3 ml 1 M Tris-HCl (pH 7.5), 300 μl 1 M $MgCl_2$, 60 μl 100 mM dGTP, 60 μl 100 mM dATP, 60 μl 100 mM dTTP, 60 μl 100 mM dCTP, 300 μl 1 M DTT, 1.5 g PEG-8000 and 300 μl 100 mM NAD.

10X Cre Recombination Buffer was prepared by mixing 3.3ml 1 M NaCl, 5 ml 1 M Tris-HCl, 1 ml 1M $MgCl_2$, adding ddH_2O to 10 ml with a final concentration of 330 mM NaCl, 500 mM Tris-HCl, 100 mM $MgCl_2$ and adjusting the pH to 7.5 at 25°C. The mixture was aliquoted to 1ml microcentrifuge tubes and stored in -20°C for longer use.

Yeast lysis buffer for yeast genome DNA extraction was composed of 10 mM Tris-HCl with pH of 8.0, 0.1 M NaCl and 1% w/v Triton X-100. RNase A was prepared with the stock concentration of 25 mg/mL in 50% v/v glycerol. The final working concentration was 200-250 μg/mL.

Frozen competent cell (FCC) solution for making yeast competent cells was composed of 5% (v/v) sterile glycerol and 10% (v/v) sterile DMSO diluted with sterile ddH_2O.

Table 2.2 List of commercial kits and reagents for DNA related experiments

Name of reagent	**Vendor**
QIAprep Spin Miniprep Kit	QIAGEN
QIAquick Gel Extraction Kit	QIAGEN
PureLink® PCR Purification Kit	Thermo Fisher Scientific
96-well DNA purification plate	BIO BASIC SD5007
Zymoprep™ Yeast Plasmid Miniprep II	Zymo research
Phusion® High-Fidelity DNA polymerase	NEB
Q5® High-Fidelity DNA Polymerase	NEB
KAPA HiFi PCR Kit	KAPABIOSYSTEMS
GoTaq® Green Master Mix	Promega
PrimeSTAR Max DNA Polymerase	TaKaRa
Restriction enzymes	NEB
T4 DNA ligase 2,000,000 units/ml	NEB
Zero Blunt® TOPO®	Invitrogen
CloneJET PCR Cloning Kit	Thermo Fisher Scientific
2-Log DNA Ladder (0.1-10.0 kb)	NEB
TopVision Agarose	Thermo Fisher Scientific
6X DNA gel loading buffer	NEB
BigDye® Direct Cycle Sequencing Kit	Thermo Fisher Scientific
Invitrogen™ SYBR™ Safe DNA Gel Stain	Thermo Fisher Scientific
PEG 3350	Sigma-Aldrich
Dimethyl Sulfoxide (DMSO)	Sigma-Aldrich
Herring Sperm DNA	Promega
Lithium acetate	Sigma-Aldrich
$CaCl_2$	Sigma-Aldrich
0.5mm Zirconia/Silica Beads	BioSpec Products
Phenol:Chloroform:Isoamyl alcohol	Sigma-Aldrich
5M Betaine solution	Sigma-Aldrich

2.2.3.2 Protein purification related reagents

The commercial protein purification related reagents are listed in Table 2.3. The remaining buffers were prepared as follows.

HisA buffer was composed of 50 mM Tris-HCl pH 8.0, 50 mM imidazole, 500 mM NaCl. It was filter-sterilized with 0.22 µm filter and degassed. HisB buffer was composed of 50 mM Tris-HCl pH 8.0, 500 mM imidazole, 500 mM NaCl. The difference between HisA and HisB lies in imidazole concentration.

Resolving gel buffer (16 ml of 15% buffer) was made with 3.7 ml ddH_2O, 8 ml 30% acrylamide/bis solution, 4 ml 1.5 M Tris pH 8.8, 160 µl 10% SDS, 160 µl 10 % APS, 16 µl TEMED. Stacking gel buffer (10 ml of 6% buffer) was made with 5.3 ml ddH_2O, 2 ml 30% acrylamide/bis solution, 2.5 ml 0.5 M Tris pH 6.8, 100 µl 10% SDS, 100 µl 10% APS, 10 µl TEMED.

10X SDS-PAGE running buffer was made with 300 g Tris base, 1440 g glycine and 100 g SDS and dissolved in 10 L ddH_2O (in fume hood) and the pH was adjusted to 8.3. The 1X running buffer was diluted from the 10X running buffer to working concentration. 5X SDS-PAGE loading buffer (10 ml) was composed of 250 mM Tris-HCl pH 6.8, 10% SDS, 30% glycerol, 5% β-mercaptoethanol and 0.02% bromophenol blue. 1% Bromophenol blue solution was prepared by dissolving 100 mg bromophenol blue powder in a 15 ml falcon tube in 10 ml of ddH_2O. The solution was covered with foil for storage.

Protein storage buffer was composed of 15 mM Tris-HCl, 250 mM NaCl, 50% Glycerol. The pH was adjusted to 8.0 at 25°C.

1M IPTG was prepared by dissolving 2.38 g IPTG in 10 ml ddH_2O and filter sterilized through a 0.22µm syringe filter. The solution was stored at -20°C.

Table 2.3 List of protein purification related reagents

Name of reagent	**Vendor**
30 % acrylamide/Bis solution (29.2/0.8%)	Severn Biotech Ltd.
SDS: 10% solution in water	Sigma-Aldrich
HisTrap FF column	GE Healthcare
TEMED - Tetramethylethylenediamine	Sigma-Aldrich

Page ruler plus prestained protein ladder	Thermo Fisher Scientific
0.45 µm syringe filter	Millipore
Liquid nitrogen	University store
Cellulose dialysis tubing 21.3 mm diameter 8 kDa molecular weight cutoff	Biodesign
Dialysis clips	Clas Olsen
50 ml polycarbonate centrifuge tubes	Thermo Fisher Scientific

2.2.3.3 Yeast protein extraction and Western blot reagents

The commercial reagents for western blotting are listed in Table 2.4. The remaining buffers were prepared as follows. The SDS-PAGE Sample Buffer (10 ml) was composed of 0.06M Tris-HCl pH 6.8, 2% (w/v) SDS, and 5% (v/v) 2-mercaptoethanol. The PBST buffer was made by adding 0.05% Tween-20 in PBS buffer. The blocking buffer was made by mixing 5% milk powder in PBST solution.

Table 2.4 Kits and reagents for western blot

Name of reagent	**Vendor**
4–15% Mini-PROTEAN™ TGX Stain-Free™ Protein Gels	Bio-Rad
Trans-Blot® Turbo™ Mini Nitrocellulose Transfer Packs	Bio-Rad
Western Bright Chemipen	labtech
Restore™ Western Blot Stripping Buffer	Thermo Fisher Scientific
Coomassie™ Dye Protein Gel Stains	Thermo Fisher Scientific
Trans-Blot® Turbo™ Mini Nitrocellulose Transfer Packs	Bio-Rad
Carestream X-Ray film Kodak MXB Blue sensitive 18x24 cm	Kodak
Trans-Blot® Turbo™ Transfer System	Bio-Rad
Precision Plus Protein™ WesternC™ Blotting Standards	Bio-Rad
Clarity Western ECL Substrate	Bio-Rad

Tween-20	Sigma-Aldrich
Monoclonal ANTI-FLAG® M2 antibody produced in mouse	Sigma-Aldrich
Anti-Mouse IgG (whole molecule)–Peroxidase antibody produced in rabbit	Sigma-Aldrich
PBS tablet	Sigma-Aldrich
Skim Milk Powder	Sigma-Aldrich
Hypercassette Autoradiography Cassettes	Amersham

2.2.3.4 RNA extraction reagents and RT-PCR kit

The reagents for yeast RNA extraction include proteinase K buffer, precipitation solution and wash buffer solution. The components of the buffers are listed in Tables 2.5, 2.6 and 2.7 respectively. The precipitation solution was stored at 4°C. For RNA reverse transcription, the SuperScript® III First-Strand Synthesis SuperMix was purchased from Thermo Fisher Scientific. RNAse AWAY® Surface Decontaminant was purchased from VWR to avoid RNA degradation by RNAse contamination.

Table 2.5 Proteinase K Buffer solution (2ml)

Reagent	Stock concentration	Final concentration	Volume
Tris, pH 8.0	1 M (100X)	10 mM	20 µl
EDTA, pH 8.0	0.5 M (100X)	5 mM	20 µl
NaCl	4 M	150 mM	75 µl
SDS	10% w/v (10X)	1% v/v	200 µl
Proteinase K	20 mg/mL (50X)	0.4 mg/mL	40 µl
ddH_2O			1645 µl

Table 2.6 Precipitation solution (2ml)

Reagent	Stock concentration	Final concentration	Volume
Potassium acetate	7 M	3 M	857 µl
Glacial acetic acid	100%	11.5% v/v	230 µl
ddH_2O			913 µl
Total			2000 µl

Table 2.7 Wash Buffer solution setup (4ml)

Reagent	Stock concentration	Final concentration	Volume
Tris-HCl, pH 7.5	1 M (100X)	10 mM	40 µl
Potassium acetate	7 M (100X)	60 mM	34.2 µl
EtOH	100%	60% v/v	2400 µl
ddH_2O			1525.8 µl
Total			4000 µl

2.2.3.5 Genetic inducers

Two genetic inducers were used in the study. β-Estradiol (E2257, powder, γ-irradiated, BioXtra, suitable for cell culture) was purchased from Sigma-Aldrich. The 1000X stock solution (1mM) was made by dissolving 1mg in 3.67 ml ethanol and stored at -20°C. Mifepristone/RU486 (M8046) was also purchased from Sigma-Aldrich. A 30mM stock solution was first made by dissolving 100 mg in 7.76 ml ethanol and stored at -20°C. The 1000X stock solution (1 mM) was further diluted from the 30mM solution for storage.

2.2.3.6 Chemical extraction and purification

The reagents and materials for extracting either violacein or carotenoids are listed in Table 2.8.

Table 2.8 List of reagents and materials for chemical extraction from yeast

Name of reagent	Vendor
Pyrogallol	Sigma-Aldrich
Hexane	Sigma-Aldrich
Violacein from *Janthinobacterium lividum*	Sigma-Aldrich
β-carotene	Sigma-Aldrich
PP screw caps, ND10 with septum	VWR
Screw neck vials, ND10	VWR
Whatman Puradisc Syringe Filters 4 mm 0.2 µm	GE life sciences

Sterile cotton swab	VWR
Microlance Hypodermic Needle 21G x 1.5"	BD Biosciences
Oral Syringes 1ml	BD Biosciences

2.2.4 Equipment

The key equipment used in the study is listed in Table 2.9.

Table 2.9 List of key equipment

Name of equipment	**Vendor**
ProFlex PCR System	Thermo Fisher Scientific
Infinite® 200 PRO multimode microplate readers	Tecan
ÄKTA chromatography system	GE Healthcare
Attune@ NxT Acoustic Focusing Cytometer	Life Technologies™
BD FACSAria III Cell Sorter	BD Biosciences
MSE Soniprep 150 Plus Ultrasonic Disintegrator	MSE
NanoDrop™ 2000 Spectrophotometer	Thermo Fisher Scientific
Gel Doc™ XR+ Imager	Bio-Rad
MiniStar centrifuge	VWR
Megafuge 8	Thermo Fisher Scientific
Eppendorf® Vacufuge® Concentrator	VWR
Agilent 1100 series HPLC system	Agilent
Zeiss Axio Imager microscope	ZEISS

2.3 DNA cloning methods

2.3.1 Gene synthesis and amplification

2.3.1.1 Build a Genome protocol

- Design of oligonucleotides

The online software GeneDesign (http://54.235.254.95/gd/) was used for designing 20 oligonucleotides that were around 60 nucleotides long each for the construction of the 1Kb Dre recombinase coding sequence.

- Primer preparation

The purchased oligonucleotides were diluted to a working stock of 6 µM. To make the templateless primer mix (TPM), 10 µl of each of the 6 µM oligonucleotides were mixed together to a final volume of 200 µl. The final concentration of each primer was 300 nM. To make an outer primer mix (OPM), 25 µl each of the first and last primers were mixed together and the final concentration of each primer was 3 µM.

Table 2.10 Reaction setup for templateless PCR (25 µl)

Reagent	Stock Concentration	Final Concentration	Vol/25 µl reaction
dNTPs	2.5 mM	200 µM	2.0 µl
TPM	300 nM	30nM	2.5
Buffer	10x	1x	2.5 µl
Phu	-	-	0.5 µl
H_2O	-	-	17.5 µl

Table 2.11 PCR program for templateless PCR

Program	PCR temperature condition (1.5-2hs)
1	94°C, 3 min
2	55°C, 30 s
3	72°C, 1 min
4 (5 cycles)	94°C, 30 s
	69°C, 30 s
	72°C, 1 min
5 (5 cycles)	94°C, 30 s
	65°C, 30 s
	72°C, 1 min
6 (20 cycles)	94°C, 30 s
	61°C, 30 s
	72°C, 1 min
7	72°C, 3 min
8	10°C, forever

- Templateless PCR

There are two PCR steps for the assembly from oligonucleotides. The first PCR reaction was a templateless PCR. The reaction was first set up as in Table 2.10. The reaction was vortexed briefly and 22.5 µl were aliquoted per tube. To each tube, 2.5 µl of TPM (300nM templateless primer mix) or positive control were added for a final volume of 25 µl. The PCR program is listed in Table 2.11 and after the templateless PCR, PCR products were tested on a gel by electrophoresis. Templateless PCR product was diluted 1/5 as template for 'Finish PCR'.

- Finish PCR

The second part of the PCR is called 'Finish PCR'. The reaction was set up as in Table 2.12. The reaction was vortexed briefly and 20.5 µl were aliquoted per tube. To each tube 2.0 µl of OPM (3 µM outer primer mix) for the Finish PCRs or 2.0 µl of water for the control reaction was added to the reaction. Finally, 2.5 µl of the 1/5 diluted templateless PCR product was used as template for the Finish PCR. The PCR program is shown in Table 2.13. PCR products were tested on a gel by electrophoresis.

Table 2.12 Reaction setup for Finish PCR (25µl)

Reagent	**Stock Concentration**	**Final Concentration**	**Vol/25 µl reaction**
dNTPs	2.5 mM	200 µM	2.0 µl
OPM	3 µM	240 nM	2.0 µl
templateless PCR product	NA	NA	2.5 µl
Buffer	10x	1x	2.5 µl
Phu	-	-	0.5 µl
H_2O	-	-	15.5 µl

Table 2.13 PCR program for finish PCR

Program	PCR Reaction condition (about 1.5hs)
1	94°C, 3 min
2	55°C, 30 s
3	72°C, 1 min
4 (25 cycles)	94°C, 30 s
	55°C, 30 s
	72°C, 1 min
5	72°C, 3 min
6	10°C, forever

2.3.1.2 PCR amplification

For DNA fragments that can be amplified from available templates, either Phusion, Q5, PrimeStar or KAPA polymerase PCR systems were used for amplification. Phusion was generally used for gene amplification; Q5 was used instead if non-specific bands could not be avoided by Phusion amplification; PrimeStar premix was used for high throughput DNA amplification in the SynNP project; KAPA was used for difficult amplifications including high GC or long fragment PCR (>5 Kb) and PCRs using primers with inevitable secondary structures like the recombination sites. Taq polymerase was used for verification tests including bacterial and yeast colony PCR. Manufacturers' instructions for each PCR method were found online and followed for different experiments.

2.3.1.3 Commercial DNA synthesis

DNA oligonucleotides of fewer than 79 bp length were all purchased from Integrated DNA Technologies (IDT) (25nmol if below 60bp and 4nmol ultrapure DNA if above 60 bp). Double stranded DNA samples were purchased from either IDT as standard gBlocks, ThermoFisher as GeneArt strings, Twist Biosciences as insert in pUC19 vector or Gen9 as insert in Gen9 vec0002.

2.3.2 DNA assembly methods

2.3.2.1 Blunt end cloning

- TOPO cloning

The assembly reaction was set up as listed in Table 2.14 with the Zero Blunt® TOPO® cloning kit. The reaction was gently mixed and incubated for 10 min at room temperature (22–23°C). For DNA more than 1kb, the incubation time was extended to 20-30 min at room temperature.

Table 2.14 Reaction setup for TOPO cloning

Reagent	**Volume(µl)**
Fresh PCR product	0.5-4
Salt Solution	1
Water	Add to total volume of 5µl
PCR™II-Blunt-TOPO®	1
Final Volume	6

- CloneJET PCR Cloning

The assembly reaction was set up as listed in Table 2.15. The ligation mixture was incubated at room temperature (22°C) for 5 min.

Table 2.15 Reaction setup for CloneJET cloning

Reagent	**Volume(µl)**
2X Reaction Buffer	10
Fresh PCR product	1
Water	Up to 19µl
pJET1.2/blunt Cloning Vector (50 ng/µl)	1
T4 DNA Ligase	1
Final Volume	20

2.3.2.2 Restriction enzyme cloning

DNA fragments were first digested by restriction enzymes (NEB) and bands with expected lengths were gel-extracted and purified using the QIAquick Gel Extraction Kit. The concentration of purified DNA was measured by NanoDrop™ 2000 Spectrophotometer. The insert DNA and vector DNA were added to the T4 ligation

system with a molar ratio of 3:1. Then 2 µl 10X T4 ligase buffer, 1 µl T4 DNA ligase and ddH_2O were added to the reaction with a final volume of 20 µl. The reaction was incubated at 25°C for 2h. Transformation was done directly after or stored in -20°C for next day use.

2.3.2.3 Golden Gate assembly

- BsaI/Esp3I-based Golden Gate Assembly

In a 20 µl reaction, 1 µl BsaI, 1 µl T4 ligase, 2 µl 10x T4 ligase buffer, 0.2 µl 100X BSA, and equal molar DNA fragments (eg, purified PCR products) with a total quantity of 500ng were mixed and the reaction was made up to 20 µl with ddH_2O. The following program was used for incubation in a thermocycler: 10-30 cycles of (37°C for 5 min, 16°C for 10 min), then 50°C for 10 min, 80°C for 5 min. Transformation was done directly after or the reaction was stored at -20°C for next day use.

- BsmBI-based Golden Gate Assembly

In a 10 µl reaction, 1 µl 10 ×T4 ligase Buffer (NEB), 0.2 µl 100X BSA, 0.5 µl BsmBI, and equal-molar quantities of DNA fragments with a total quantity of 500 ng were mixed and the reaction was made up to 10 µl with ddH_2O. The following program was used for incubation in a thermocycler: 55°C for 90min, add 0.5 µl T4 DNA ligase, 10-30 cycles of (37°C for 5 min, 16°C for 10 min), then 55°C for 10 min, 80°C for 5 min. Transformation was done directly after or the reaction was stored at -20°C for next day use.

2.3.2.4 Gibson assembly

15 µl Gibson master mix (4 µl 5x Isothermal buffer((25% PEG-8000, 500 mM Tris-HCl pH 7.5, 50 mM $MgCl_2$, 50 mM DTT, 1 mM each of the four dNTPs and 5 mM NAD), 0.16 µl T5 Exonuclease (1/20), 0.25 µl Phusion Polymerase, 2 µl of 40(U/µl) *Taq* DNA ligase, 8.59 µl H_2O) was mixed with equimolar purified DNA fragments (~50 to 100 ng) to a final volume of 20 µl(Gibson et al., 2009). The reaction was incubated at 50 °C for 1 h. Transformation was done directly after or the reaction was stored at -20°C for next day use.

2.3.2.5 *In vivo* yeast homologous recombination assembly

Linear DNA fragments purified by either gel extraction purification kit or PCR purification kit were used for yeast homologous assembly. 500ng of each fragment were transformed by frozen yeast transformation protocol. The transformants were plated on corresponding selective media and cultured at 30°C for 2-4 days. Following plasmid extraction, gDNA extraction or yeast colony PCR were performed for screening clones for correct assembly.

2.3.3 Assembly screening and verification

2.3.3.1 Direct screening

- Bacterial cloning

Antibiotic selection. Typical antibiotic resistance genes *AmpR* and *KanR* were used for plasmid selection. Different antibiotics were used alternatively in the SynNP project to achieve hierarchical construction (see 2.2.1 for concentrations).

Color selection. Bacteria colonies expressing RFP have rose-carmine color after a certain time of culture. In order to accelerate the speed of color development, bacteria with the RFP gene were incubated at 4°C after colonies had formed at 37°C. For clones with insert containing RFP, red colonies were picked for verification; for clones with RFP as counter-selection marker, white colonies were picked for verification.

Counter-selection with *ccdB* toxic gene. The *ccdB* gene was included in several construction vectors and was used for counter-selection. Strains with self-ligated vectors did not survive and colonies were picked directly from the plates for next step verification. This is a more efficient counter-selection marker replacing the RFP gene.

- Yeast cloning

Auxotrophic selection. Commonly used auxotrophic markers in yeast including *HIS3, LEU2, URA3, LYS2, TRP1* and *ADE1* were used in this study. Based on genotype and selection marker of the clone, single or multiple amino acid dropout media were used for selection (see 2.2.2).

Antibiotic selection. Clones with antibiotic resistance gene *KanMX* were selected in Geneticin (G418) containing media.

5-FOA counter-selection. The expression of *URA3* can convert the substrate 5-FOA (5-Fluoroorotic acid) to a toxic form, 5-fluorouracil, that can lead to cell death. Clones requiring the removal of *URA3* were selected on 5-FOA-containing medium.

2.3.3.2 Colony PCR verification

- Bacterial colony PCR

Bacterial colony PCR was a routine method for preliminary screening of DNA cloning in bacteria. The PCR was performed with GoTaq Green master mix and the reactions were set up as shown in Table 2.16. Bacterial colonies were picked from culture plates using sterile tips and were dipped into the PCR reaction as template and selection medium for strain preservation. The volume of the reaction was adjusted to 6μL, 12.5μL or 25μL in different experiments.

Table 2.16 Reaction setup for GoTaq Green bacterial colony PCR

Reagent	**Volume (μl)**
2x GoTaq Green mix	12.5
ddH_2O	10
10 μM Forward primer	1.25
10 μM Reverse primer	1.25

- Yeast colony PCR

Yeast colony PCR is similar to bacterial colony PCR. The difference is simply that boiling yeast will not efficiently release the DNA template for PCR. Therefore, yeast was pretreated with 20 mM NaOH and boiled at 95°C for 10 min to be used as template. Moreover, the annealing time was extended to 2 min/kb in yeast colony PCR. Specifically, a 96-well plate yeast colony PCR was performed for the SynII SCRaMbLE PCR tag assay. The procedure was performed as follows. A sterile pippet tip was used to pick a yeast colony and handle it to the master 96-well plate in one well containing 50 μl sterile (nuclease-free) ddH_2O. The cells were re-suspended completely by pipetting up and down. 25 μl freshly-prepared 40 mM NaOH was aliquoted to a corresponding number of PCR tubes or wells of a PCR plate. 25 μl of the resuspended cells was mixed with 25μl 40 mM NaOH (20 mM NaOH final conc).

The plate was put in the PCR machine for heating at 95°C for 10 min, then cooled to 4°C. The GoTaq Green PCR reaction mixture was prepared for 100 PCR reactions and the set up for each reaction was as shown in Table 2.17. After mixing by vortex, 10 μL of the mixture was aliquoted to each well of the plate and 2.5 μL NaOH-treated cell lysate was added to each well to a final volume of 12.5 μL. The PCR program is shown in Table 2.18. 1.3% agarose gel electrophoresis was used for testing these short DNA bands (<500bp).

Table 2.17 Reaction setup of yeast Colony PCR

Reagent	**Volume (μL)**
2x GoTaq Green mix	6.25
ddH_2O	2.25
10 μM Forward primer	0.75
10 μM Reverse primer	0.75

Table 2.18 Special program for PCR-tagging for yeast synthetic strains

Step	**Temperature (°C)**	**Time**
1	95	4 min
2	94	10 s
3	52	1.5 min
4	72	1 min
5 (cycle number)	30X	Step 2-4
6	72	5 min
7	4	Hold

2.3.3.3 Plasmid DNA miniprep and digestion

For Single tube miniprep, the experiment was performed using the QIAprep Spin Miniprep Kit according to the manufacturer's instructions.

For 96-well plate miniprep, some modifications were made from the original protocol. The bacteria were cultured in a sterile 96-well U bottom deep well plate. 1.2 ml LB medium with antibiotics was added to each well and the plate was sealed

with breathable membrane. The cells were cultured at 37°C and a shaking speed of 300 r.p.m for 20-24 h. The culture plate was spun down at 5000 g for 5 min and the supernant was discarded. The cells were first re-suspended with 200 μl P1 buffer. Then 200 μl P2 lysis buffer was added and the plate was tilted carefully as far as possible to mix the cells and buffers. The mixture was incubated for 1 or 2 min for cell lysis. Then 280 μL of N3 neutralization buffer was added and the plate was carefully tilted again as far as it could be to mix the buffer and generate the white floc. The reaction plate was centrifuged at 5000 g for 20-30 min and the supernant was transferred to a new DNA purification column plate. The culture plate was used as the waste collection plate and the column plate was put on top of it. The plates were centrifuged at 5000 g for 3 min. Then 500 μl PB buffer was added to each well to get rid of remaining protein. The plates were centrifuged again at 5000 g for 3 min. The waste liquids were discarded. Then 750 μL PE buffer was added and centrifuged at 5000 g for 5 min as a final wash. The column plate was then transferred on top of a 96-well PCR plate. 50 μl 50°C pre-warmed ddH_2O was added to each well of the column. After 1-2 min incubation, the DNA was eluted to the PCR plate by centrifuging at 5000 g for 3 min. DNA concentration and quality was evaluated by NanoDrop.

2.3.3.4 Yeast genomic DNA extraction

5 ml of yeast cells from overnight liquid culture was centrifuged at 17000 g for 1 min and the supernatant was discarded. The pellet was re-suspended in 400 μl of lysis buffer (10 mM Tris-HCl pH 8.0, 0.1 M NaCl, 1% w/v Triton X-100). 200 μl of 0.5 mm glass beads and 400 μl of phenol:chloroform:isoamyl alcohol (25:24:1 v/v) were added to each tube in the fume hood and the tubes were vortexed for 10 min. The mixture was centrifuged at 17000 g for 10 min and the aqueous phase was carefully transferred into a new 1.5 ml microcentrifuge tube. Then 750 μl of cold absolute ethanol was added to the tube and mixed by vigorously shaking and the mixture was put in the -20°C freezer for 10 min. The mixture was centrifuged at 17000g for 10 min and the supernatant was discarded. The pellet was washed with 300 μl of 70% v/v

ethanol at room temperature and then centrifuged at 17000 g for 5 min. The supernatant was discarded and the pellet was dried at 45°C in the vacufuge concentrator for 10-15 min to get rid of the remaining ethanol. The pellet was re-suspended in 100 μl of EB buffer (10 mM Tris-HCl pH 8.5). To remove the residual RNA 1 μl of RNase A (stock: 25 mg/ml in 50% v/v glycerol) was added and the mixture was incubated for 10 min at 60°C. Then 200 μl of TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0) was added for removal of phenol. An equal volume of chloroform:isoamyl alcohol (25:1 v/v ratio) was added and the mixture was vortexed for 3 min. The mixture was centrifuged at 17000 g for 5 min and the upper layer was collected and transferred to a new tube. For final DNA precipitation, 500 μl of cold absolute ethanol and 20 μl of 3 M sodium acetate (pH 5.2 – 10:1 v/v layer:salt ratio) was added and the mixture was incubated in the -20°C freezer for 10 min. The mixture was centrifuged at 17000g for 10 min and the supernatant was discarded. The pellet was washed with 300 μL of 70% v/v ethanol at room temperature, centrifuged again at 17000 g for 10 min with the supernatant discarded. Finally, the pellet was dissolved in 50 μl of ddH_2O.

2.3.3.5 BigDye DNA sequencing reaction

For basic Sanger sequencing of DNA constructs, a BigDye sequencing reaction was set up as in Table 2.19 and performed with the PCR program listed in Table 2.20. After the PCR reactions were finished, the samples were sent to Edinburgh Genomics for sequencing.

Table 2.19 Reaction setup for BigDye sequencing PCR

Reagent	**Volume (μL)**
5X Sequencing Buffer	2
BigDye	1
10 μM sequencing primer	0.32
DNA	100-200ng plasmid DNA 10ng for PCR DNA products
ddH_2O	Add up to 10 μL

Table 2.20 Program of BigDye sequencing PCR

Step	Temperature (°C)	Time
1	95	30 s
2	55	20 s
3	60	4 min
4 (cycle number)	32X	Step 1-3
5	4	Hold

2.3.4 Transformation and strain storage

2.3.4.1 Bacteria transformation

- Competent cell preparation

5 ml bacteria source strain (DH5α, BL21 or ccdB Survival™ 2) was cultured in LB overnight. Then all the 5 ml culture was put in 500 ml SOB media early in the morning and grown in 37°C for 2-2.5 h to OD_{600} 0.4. A box of ice was prepared for keeping the cells cold during washing with CCMB buffer (see 2.2.1) and the standing centrifuge was precooled to 4°C. The culture was first centrifuged at 4°C at 4000 g for 5 min and re-suspended in 100 ml ice-cold CCMB. The centrifuging procedure was repeated for three more times and each time the cells were washed with 50 ml, 50 ml and a final volume of 20 ml CCMB. The cell suspension was further incubated on ice in the 4°C cold room for 30 min-2 h. The cells were then aliquoted to 1.5 ml microcentrifuge tubes and stored at -80°C. The transformation efficiency was tested with 20 pg pUC19 plasmid before actual use.

- Procedure for normal transformation in 1.5ml microcentrifuge tube

The water bath was first equilibrated to 42°C. The SOC. medium and LB plates containing corresponding antibiotics were warmed to room temperature. The competent cells were thawed on ice and aliquoted to 50 µl for each transformation. 10ng plasmid or half the volume of the cloning reaction was added into 50 µl competent cells and the mixture was incubated on ice for 20-30 min. After the incubation, the cells were heat-shocked for 45 s at 42°C in the water bath (for DNA

larger than 15kb, heat-shock time was increased to 75 s). The tubes were then immediately transferred to ice and kept on ice for 2-5 min. The cells were then added to 500 µL SOC. medium or 800 µl LB medium for recovery at 37°C for 1 h. The cells were later centrifuged at 17000 g for 10 s and re-suspended in 250 µl medium for plating. 20 µl and 200µl of cells from each transformation was plated on corresponding selective plates and incubated overnight at 37°C.

- Procedure for transformation in 96-well plate

The procedure for plate transformation was very similar to the one for single tubes with only a few modifications. First, LB OmniTray plates containing corresponding antibiotics were used instead of circular plates. Second, 25 µl of competent cells were aliquoted for transformation and up to 5 µL assembly reaction was added. Third, the 96-well PCR machine was used for heat-shock and 200 µl SOC medium for each well was used for strain recovery. Instead of spreading one transformation to one plate, 12 transformations were dripped on to each single OmniTray plate. With a multichannel pipette, 30 µl of each culture was dripped on an OmniTray medium plate from one end to the other (12 samples/plate) by tilting the plate, dispensing the culture and slowly dripping the cells down the agar surface. After dripping, the OmniTray was dried in a tilted position and later incubated at 37°C overnight.

2.3.4.2 Yeast transformation

- General Yeast Transformation

A yeast overnight culture was re-inoculated to fresh medium with OD_{600} adjusted to 0.1 and cultured for around 4 h in YPD medium or around 6 h in dropout medium till the OD_{600} reached to 0.5-1.0. The cells were then harvested at 3000 g for 5 min, washed once with 10 ml of sterilized water, centrifuged for a second wash with 10ml of 0.1M LiOAc and centrifuged again. The LiOAc was poured off but the small volume remaining at the bottom of the tube was used to re-suspend the cells. At the same time, Salmon Sperm DNA (HS DNA) was boiled at 100°C for 10 min and set on ice for making the transformation reaction buffer. For one reaction, 240 µl PEG (44%), 36 µl LiOAc (1 M), 25 µl HS DNA (10 mg/ml), 5 µl target DNA (100-500 ng) and 19 µl ddH_2O

were mixed well by vortexing and then 50 µl competent cells were added on top. The mixture was first incubated at 30°C for 30 min, then 36 µl of DMSO was added and mixed immediately by inverting before another incubation at 42°C for 15-30 min. Then the cells were centrifuged at 3000 g for 1 min, re-suspended in 400 µl 5 mM $CaCl_2$ and placed at room temperature for 10 min. For auxotrophic selection, 200 µl cells were spread directly to selection plates; for antibiotic selection, the cells were further recovered in YPD for 2-3 h and then spread on an appropriate selection plate.

- Frozen yeast transformation

The procedures for making frozen yeast competent cells and the transformation were done according to instructions published by Gietz et al (Gietz & Schiestl, 2007). In this study, the yeast strain BY4741 was prepared as frozen competent cells for routine use. BY4741 was inoculated into 2 tubes of 10 ml of liquid 2 × YPAD medium and cultured overnight on a shaker at 200 r.p.m at 30 °C. A bottle of 500ml 2 ×YPAD and one sterile 2 L flask were pre-warmed in the incubator as well. After 12-16 h of growth, the titer of the yeast culture was determined with a spectrophotometer. Cell concentration in the original culture was roughly calculated with the dilution factor and OD_{600} (a suspension containing 1×10^6 cells /ml will give an OD_{600} of 0.1). Then $2.5X10^9$ cells were added to 500 ml of the pre-warmed 2× YPAD in the pre-warmed flask. The cells were cultured in the shaking incubator at 30°C until the cell titer reached at least 2×10^7cells /ml (OD: ~2.0), which normally took about 4 h. The cells were harvested by centrifugation at 3000g for 5 min, washed with 50ml sterile water, spun down again and re-suspended in 5 ml sterile water. The cell suspension was transferred to a suitable sterile centrifuge tube and spun down again at 3000 g for 5 min at room temperature. The cell pellet was re-suspended in 5 ml filter sterilized frozen competent cell (FCC) solution (5% v/v glycerol, 10% v/v DMSO). The samples were finally dispensed in 50 µl aliquots into an appropriate number of 1.5 ml microcentrifuge tubes. The tubes were then placed into a Styrofoam rack with lid. The rack with tubes was placed upright in a larger box (Styrofoam or cardboard) with additional insulation such as Styrofoam chips or newspaper to reduce the air space around the sample box. This helped the samples to freeze slowly, which is essential

for good survival rates. The large Styrofoam container was put in a -80°C freezer overnight. The Styrofoam rack containing the frozen yeast cells was removed from the freezing container and stored at -80°C.

The transformation procedure was slightly different from the normal transformation. The cell samples were first thawed in a 37°C water bath for 15–30 s and spun down at 13000 g for 2 min to remove the supernatant. The HS DNA was at the same time boiled at 100°C for 5 min and chilled on ice immediately. A 360 µl transformation reaction was set up for each sample as listed in Table 2.21 and was used to re-suspend the cell pellet by vortexing. The cell suspension was incubated in a 42°C water bath for 20–60 min. After the incubation, the cells were spun down at 13000 g for 30 s and the supernant was discarded. The cells were re-suspended in either 1 ml of sterile water (prototrophic) or YPD (antibiotics). For auxotrophic selection, 200 µl cells was spread directly onto selection plates; for antibiotic selection, a further recovery in YPD at 30°C for 2-3 h was performed before plating the cells on a suitable selection plate.

Table 2.21 Reaction set up for frozen yeast transformation

Transformation mix components	**Volume(µl)**
PEG 3350 (50% (w/v))	260
LiOAc 1.0 M	36
single-stranded carrier DNA (10.0 mg ml^{-1})	10
Plasmid DNA plus sterile water	54
Total volume	360

2.3.4.3 Strain storage and strain database

Equal volume of 30% autoclaved glycerol and either bacteria or yeast cell culture growing in log phase were mixed well and stored in -80°C in cryogenic tubes. The information on stored bacterial or yeast strains was uploaded to the laboratory database http://database.cailab.org/.

2.4 Growth assays

2.4.1 Serial dilution and spotting assay

In a 96-well plate, 100 µl liquid culture of target strains were added in the first lane (for survival rate comparison on different plates, the OD_{600} of the samples were adjusted to 1.0 before serial dilution; for survival rate comparison between different strains on plates of the same type, the samples were spotted directly without OD_{600} adjustment). 90 µl distilled water or PBS buffer was aliquoted into the rest of the lanes. The serial dilution was done by sequentially transferring 10 µl culture to a new well with 90 µl distilled water or PBS buffer, mixing well and transferring to next well with distilled water or PBS buffer. For spotting assay, the plate was pre-warmed to 30°C. If spotting onto synthetic dropout plates, 10 µl of samples were spotted; if spotting to YP plates, 5 µl of samples were spotted. The plates were dried near a Bunsen burner and incubated at 30°C for 2-3 days.

2.4.2 Growth curves and fluorescence curves

Growth and fluorescence curves were measured by the Infinite® 200 PRO multimode microplate reader (Tecan). Greiner 96 Flat Bottom Transparent Polystyrene plates (Cat. No.: 655101) were used for general growth curves, and BD Falcon 96 Flat Bottom Black Polystyrene plate (Cat. No.: 353219) were used for fluorescence measurement. The parameters were as follows: culture temperature for yeast was 30°C; shaking (linear) duration and amplitude were 550 s and 6 mm; kinetic cycle was 180 and multiple reads were recorded for mean calculation. Cell density was measured with the absorbance wavelength of 595nm; GFP fluorescence was measured with excitation wavelength of 485 nm, emission wavelength of 535 nm and fluorescence bottom reading mode.

2.5 *In vivo* recombinase function and ligand induction assays

- Plate induction

For plating assay, the cells transformed with recombinase expressing device and function test device were replica plated from SC-His-Leu glucose plates to SC-His-Leu-Ura plates containing 1 µM of either β-estradiol or RU486. The original plate was

cultured for one more day and the β-estradiol or RU486 induction plate was cultured for 2-3 days at 30°C. For the spotting assay, the target strain was cultured overnight in non-hormone containing liquid medium. The overnight culture was adjusted to OD_{600} 1.0 and spotted on hormone-containing plates and non-hormone containing plates for comparison.

- Liquid induction

The target strain was cultured overnight in non-hormone containing liquid medium. Each overnight culture was re-inoculated to two tubes of fresh media and the OD_{600} of the samples were adjusted to 0.1. In one tube, the corresponding hormone inducer was added and in the other tube an equal volume of ethanol was added as control. The spotting assay was done for both testing and control strains at different time points on SC-His-Leu and SC-His-Leu-Ura glucose plates. The plates were cultured for 2-3 days at 30°C for growth pattern comparison.

2.6 Yeast chromosome directed evolution assay

- SCRaMbLE

Synthetic strains with Cre expression plasmid were cultured in SC-His liquid medium overnight. For liquid induction, all the strains were re-inoculated to OD_{600} of 0.1 in two tubes in 10 ml fresh SC-His liquid medium; 10 µl 1 mM β-estradiol/RU486 was added to one of the tube and 10 µl ethanol was added to the other tube as control. The induced and non-induced strains were serially diluted and spotted on dropout plates after 24 h and 48 h. For plate induction, after culturing at 30°C for 24 h, the samples were serially diluted and spotted on 1 µM inducer-containing plates and kept at 30°C for another 48 h.

- SCRaMbLE-in

Synthetic strain *SynII* was first transformed with a SCRaMbLE-in device and selected on an SC-Leu plate. The SCRaMbLE-in device carrying *SynII* was further transformed with the Cre expression device (either P_{SCW11}CreEBD or P_{TDH3}CreEBD) or the control plasmid pRS413 and selected on a SC-His-Leu plate. After 3 days of 30°C incubation, single isolates were picked for SCRaMbLE-in induction. A single isolate was inoculated

to SC-His-Leu liquid medium and cultured at 30°C overnight. The cultures were re-inoculated to OD_{600}=0.1 in 1μM β-estradiol containing SC-His-Leu liquid medium for 24h and 48h induction. After the induction, cells were serially diluted and spotted on SC-Leu and SC-Leu 5-FOA plates for integration efficiency comparison. The cells were also spread on large SC-Leu 5-FOA plates with proper dilution for single colony isolation.

2.7 Protein expression, extraction, purification and storage

- Protein expression

The verified pET-28(a) expression plasmids for Cre, Dre, and VCre were transformed into chemically competent *E. coli* BL21 (DE3). A single colony of each transformation was inoculated to 10 ml LB media, supplemented with Kanamycin (50 μg/ml) and incubated overnight at 37°C with shaking at 180 r.p.m. 10 ml of the overnight pre-culture was then transferred into a 2.5 L baffled flask with 1 L of LB medium supplemented with kanamycin and incubated at 37 °C with shaking at 180 r.p.m. The cell growth was monitored by recording the OD_{600} of the culture at regular intervals and the recombinant protein production was induced at an OD_{600} of 0.6 with the addition of 1 mM IPTG. It usually took around 2.5 h to reach OD_{600} 0.6; the OD_{600} should not exceed 0.8 to ensure a good protein expression. The incubation temperature was then reduced to 18°C to maximize protein expression and the cells were cultured for another 16 h. An alternative induction method was culturing the cells at 37°C for 4h with induction. The efficiency of the two methods varied for different proteins.

- Protein extraction

The cells from the 1 L cultures were harvested by centrifugation at 4000 g for 20 min at 4^{o}C, and re-suspended in 10-fold (volume per gram of cell pellet) of PBS. The resuspended cells were transferred to a 50 ml falcon tube before a second centrifugation step at 4000 g for 20 min at 4^{o}C. The cell pellets were flash cooled in liquid nitrogen and stored at -80^{o}C until proceeding with the next steps. The frozen cell pellets were re-suspended in 10-times (volume/cell weight) of HisA buffer and

defrosted in a water bath at room temperature until the cell pellet was fully resuspended. The cells were lysed by sonication at 4 °C, with ten 15-second bursts of sonication at 10 μm amplitude followed by 15 s of cooling. During sonication, the cells were always kept on ice to avoid overheating and protein degradation. The cell lysate was placed into a 50 ml polycarbonate centrifuge tube and clarified by centrifugation at 35000 g for 30 min at 4^{o}C. The resulting cell extract was filtered with a 0.45 μm syringe filter into a 50 ml falcon tube. Samples from both the supernatant and the pellet were taken for SDS-PAGE to determine the level of protein in the soluble and insoluble fractions to assess protein solubility.

- Protein purification

A 5 ml HisTrap FF column was attached to an ÄKTA purification system or suitable pump system and pre-equilibrated with HisA buffer. The filtered supernatant was loaded onto the column while monitoring by the UV absorbance of the flow-through at 280 nm. When the sample was loaded on the column, it was washed with HisA buffer until the A_{280nm} returned to the baseline reading and the value remained stable for at least two column volumes. The bound his-tagged recombinase proteins were then eluted with a linear gradient of 0 – 70 % HisB buffer over 20 column volumes while collecting 3 ml fractions of the elution. As the gradient proceeded, the A_{280} measurement increased sharply at the point that the recombinase protein eluted, giving a symmetrical peak on a trace of the values.

The peak fractions were analyzed for the presence of protein at the correct molecular weight by subjecting the samples to SDS-PAGE analysis. 10 μl of each peak fraction was added to 5 μl SDS-loading buffer in a 1.5 ml microcentrifuge tube. The samples were boiled in a heat block at 98^{o}C for five min to fully denature any protein in the sample. A 15 % SDS-PAGE gel was set up in an appropriate gel-tank with the anode and cathode chamber topped up with SDS-running buffer. Any bubbles in the sample wells were removed with a pipette tip. 5 μl of pre-stained ladder was added to the first well and then the boiled samples were slowly added to other wells. The gel was run at 150 V for 1 h. The movement of the bromophenol blue loading dye and pre-stained ladder was monitored and the gel running time was adjusted to ensure

adequate separation of the bands on the ladder. The gel was disconnected from the power supply and the SDS-running buffer was drained out. The gel was removed from the tank and then carefully removed from the gel plates. The gel was transferred to a plastic container and washed five times with deionized water. The gel was stained with a Coomassie™ Dye Protein Gel Stains and pictured by Bio-Rad gel-doc to determine which fractions contain the purified recombinase.

Fractions containing target recombinase were removed from the HiTrap purification and placed in a reconstituted dialysis bag sealed with dialysis clips, or sandwich clips. The samples were dialyzed against 2000 times volume of protein storage buffer overnight at 4°C with stirring. Dre was observed to show some precipitation during the dialysis. The protein precipitate was removed by centrifugation of the protein solution at 4000 g for 30 min at 4°C. Clarified Dre was flash-cooled and stored at -80 °C as soon as possible after dialysis. The yield of VCre was much lower than that of the other two recombinases and required concentration prior to use. The extinction coefficient and molecular weight of the recombinases were obtained from the ExPASy database. The calculated molar extinction coefficient of each protein was used to estimate the concentration by measuring the A_{280} with a NanoDrop 2000 Spectrophotometer and the "Other protein (E & MW)" sample type was used for concentration measurement.

- Protein storage

For storage, 50 µl aliquots of protein at 20 µM were mixed with an equal volume of 100% glycerol in thin-walled PCR tubes, flash-cooled in liquid nitrogen and stored at -80 °C. The activity of VCre was observed to be severely compromised by slow cooling. To use recombinase from -80°C stock, the recombinases were thawed at 37°C for 10s. Repeated freeze and defrost was avoided.

2.8 *In vitro* recombinase function test assay

- Function test by digestion map variation

In a 100 µl reaction, 1 µl 10 µM recombinase, 1 µg reporter DNA with target recombination sites for testing, 10 µl 10X recombination buffer were mixed and

ddH_2O was added to make up to 100 µl. The mixture was incubated at 37°C for 2 h. The DNA after recombination was purified with the PureLink® PCR Purification Kit and digested by ScaI-HF, followed by verification by electrophoresis.

- Excision rate measurement

250 ng plasmids with recombination site flanking *RFP* cassette were mixed with 1 µl 5 µM recombinase and 5 µl 10X recombination buffer in a 50 µl reaction. The reaction was incubated at 37°C for 16 h and the recombinase was then deactivated at 75°C for 15 min. 10 µl reaction was transformed into *E. coli* competent cells and selected on LB+Ampicillin plates.

- Integration rate measurement

300ng *RFP-AmpR-ccdB* plasmid was mixed with 100 ng *KanR* plasmid. The reaction was transformed to *E. coli* competent cells and selected on LB+Kanamycin plates. Biological triplicates were performed for rate calculation. ImageJ Software was used to facilitate counting of red and white colonies.

- Promoter library integration

Twenty-five promoters from the YeastFab library were mixed and assembled with Vlox or loxJTZ15 promoter loading vector by BsmBI Golden Gate assembly. For promoter integration, the recombination reaction was set up by adding 1.5 µg promoter mixture, 800 ng HO-Vlox-CrtI Carotene acceptor circuit or HO-loxJTZ15-VioA violacein acceptor circuit, 10 µL 10X recombination buffer, VCre or Cre to a final concentration of 100 nM, and ddH_2O to 100 µL. The reaction was incubated at 37°C for 16 h, then linearized by BsmBI and transformed into BY4741 by frozen yeast transformation.

2.9 Reverse Transcription-PCR from yeast

- Yeast RNA extraction

A yeast overnight culture was re-inoculated into 20 ml yeast YPD or synthetic dropout medium to OD_{600} of 0.5. The cells were cultured for another 4-6 h until the OD_{600} reached to 3.0. The bench and pipettes were cleaned with the RNAase Away spray at first and new sterile P1000, P200 and P10 tip boxes were used specifically for the

experiment. The proteinase K buffer was pre-warmed to 65°C in a shaking heat block. The cells were collected in a 50 ml falcon tube by centrifuging at 3500 g for 2 min and the supernatant was discarded. The cell pellet was then re-suspended in 300 µl pre-warmed proteinase K buffer by gently vortex mixing to avoid excessive foaming and incubated for 15 min at 65 °C in a shaking heat block. The reaction was then chilled on ice and incubated for 20 min. 175 µl of ice-cold precipitation solution was added and vortexed for 10 s. The cell lysing mixture was spun down at 17000 g for 10 min at 4°C. The supernatant was transferred to a fresh tube and 600 µl of ice-cold absolute ethanol was added to each well. The lysate/ethanol mixture was then transferred to a binding column and spun down at 17000 g for 10 min. The flow-through was discarded. The sample was further washed with 500 µl wash buffer, spun down and the supernatant discarded. The centrifugation was repeated to remove the residual wash buffer. 100 µl ddH_2O was added into the column and incubated for 2 min and the RNA was then eluted. The RNA solution was treated with 5 µl DNase I at 37°C for 3 h to remove any DNA contamination. 250 µL ice-cold ethanol was added and the mixture was put in the freezer for 3h (or overnight as a pause point). The samples were spun at 17000 g for 20 min at 4°C. The supernatant was carefully removed. The pellet was washed with 200 µl of 70% v/v ethanol and spun down again at 17000 g for 1 min to remove residual ethanol. The pellet was dried in the vacufuge at 30°C for 30 min (covered with a breathable seal) to remove any remaining ethanol. The RNA was finally re-suspended in 60 µl ddH_2O. The concentration of the extracted RNA was measured with a NanoDrop spectrophotometer. Good quality RNA has a A260/280 value of ~2.0 and a A260/230 ratio of 2.0-2.2.

- Reverse transcription

Before reverse transcription, the extracted RNA was diluted to 1 µg/µl. The bench was cleaned again with RNAse AWAY® surface decontaminant. The SuperScript III First-Strand Synthesis SuperMix kit was used for reverse transcription (see 2.2.3.4). The reaction was set up by mixing 10 µl 2X RT Reaction Mix, 2 µl RT Enzyme Mix, 1 µl RNA (up to 1 µg) and DEPC-treated water to 20 µl. The reaction was first incubated

at 25°C for 10 min and then at 50°C for 30 min. The reaction was terminated at 85°C for 5 min and then chilled on ice. The RNA templates were removed by adding 1 µl (2U) of *E. coli* RNAse H and incubation at 37°C for 20 min. The reaction was then used as template (up to 10%) for PCR or qPCR to verify whether the cDNA was successfully reverse transcribed, and stored at -20°C until use.

2.10 Yeast protein extraction and Western blot

- Yeast protein extraction

1.5 ml yeast cells were harvested prior to stationary phase (OD_{600} = 1.0) by centrifugation at 17000 g for 1 min. The cells were first pre-treated with 200 µl of 2.0 M LiOAc for 5 min at room temperature and were centrifuged at 17000 g for 1 min, then the supernatant was removed. The cells were treated with 200 µl of 0.4 M NaOH for 5 min at room temperature. The cells were resuspended in 100 ml SDS-PAGE sample buffer and boiled for 5 min at 95°C and the lysate was centrifuged to clear cellular debris. The supernatant containing yeast whole protein extract was transferred to a fresh tube and stored at -80 °C.

- SDS-PAGE and membrane transfer

The concentration of the whole protein extraction was roughly estimated by NanoDrop. The protein samples (25-100 ng) were mixed with 2X SDS loading buffer and loaded into a Mini-PROTEAN™ TGX Stain-Free™ Protein Gel (Bio-Rad). After running at 150 V for 40 min, the gel was pictured by UV activation in a Bio-Rad Gel Doc. The proteins were transferred to nitrocellulose membrane using the Trans-Blot® Turbo™ Transfer System and the gel was imaged again after the transfer to see if there was any remaining sample protein.

- Blotting

The nitrocellulose membrane was blocked with blocking buffer at room temperature overnight at 4°C. Then the membrane was incubated with PBST-primary antibody buffer (1:2000 dilution) for 3h at 4°C. After the incubation, the membrane was washed three times with PBST for 5 min each wash. Then the membrane was incubated with PBST-secondary antibody buffer (1:5000 dilution) at room

temperature for 1h. The membrane was washed three times with PBST again, with 5 min incubation for each. The membrane was then ready for signal development.

- Clarity ECL

The two reagents in the Clarity Western ECL Substrate (Bio-Rad) were mixed in a 1:1 ratio (3.5 ml each) in a falcon tube. The membrane was removed from water and dabbed quickly on a piece of tissue and then placed on a piece of clingfilm. The ECL reagent mixture was then added on top of the membrane to cover the protein area (membrane facing upwards). The membrane was then covered with foil and left on the bench for 2 min to be developed in a dark room immediately. The Carestream X-Ray film (Kodak) was used for signal development. The developing time was adjusted for different signal intensity.

2.11 Microscopy of carotenoid-generating yeast

- DAPI staining of Yeast

This method was originally for staining the nucleus and visualization of mt-DNA since DAPI is a fluorescent stain which strongly binds to A-T rich regions in DNA. In this study, DAPI staining was used to distinguish the distribution of the yeast nucleus and the fluorescence from carotenoids produced in yeast. A DAPI/mounting solution was prepared by first dissolving 50mg p-phenylenediamine in 5 ml of PBS and adjusting the pH to ~9.0, then adding 45 ml glycerol and stirring to homogeneity (with the solution protected against the light) and adding 3 µl of DAPI (1 mg/ml). The solution was stored at either -70°C or -20°C. For cell staining, 1 ml liquid culture of log phase cells was collected by centrifugation and washed in ddH_2O. The 1 ml cell sample was re-suspended in 0.1 ml of 36% formaldehyde and incubated at RT for 1-2 h to be fixed. The cells were washed twice in 1 ml water by centrifuging at 13000 g for 5 s. The cell pellet was re-suspended in 1 ml of 70% ethanol and left at RT for 30 min for permeabilization of membranes and counter-staining. The cells were spun down at 13000 g for 5 s and re-suspended in 0.5 ml water and be sonicated for 5 s in a water bath sonicator. The cells were spun down at 13000 g for 5 s and the liquid was

discarded. 4 µl of cells were mixed with 4 µl of DAPI mounting medium for microscopy (protected from light).

- Fluorescence microscopy of carotenoids generating yeast

Microscopy was performed on a Zeiss Axio Imager microscope using objectives x100 - NA 1.40 Plan ApoChromat (oil) or x40 - NA 0.75 Plan NeuFluar (air). The filter sets of DAPI/BFP, CFP, YPF, GFP/FITC/Alexa488/Cy2, TexasRed/mCherry were used for testing the fluorescence characters of carotenoid-producing yeast. Images were captured using a Photometrics Evolve EMCCD camera operated through Micromanager software.

2.12 Flow cytometry and FACS

- 96-well fluorescence measurement

Yeast colonies were cultured in 96-well deep well plates at 30°C overnight and diluted 1:100 in fresh PBS before testing. All samples were analyzed using the Attune$^{@}$ NxT Acoustic Focusing Cytometer (Life Technologies™). A 488 nm blue laser was used for excitation and a 530/30 emission filter was used for fluorescence measurement. Cells were first gated by forward scatter (FSC) and side scatter (SSC) and 50,000 events were recorded for statistical analysis. The voltage threshold was adjusted to cover the ranges of the fluorescence. Flow cytometry data was then analyzed and visualized by FlowJo™ software 10.0.7.

- Fluorescence-activated cell sorting **(**FACS**)**

SCRaMbLEed yeast cells were cultured to log phase and diluted in PBS to a density of 10^7 - 2 X 10^7 cells/ ml. The fluorescence measurement and cell sorting were analyzed and performed by the FACSAria cell sorter (BD Biosciences). A 13 mW 488 nm coherent Sapphire laser was used for excitation and a 530/30 emission filter was used for GFP fluorescence measurement. Cells were gated by FSC-A and SSC-A and categorized to visible particles, intact cells and singlets. With further gating by Alexa Fluor 488-A, cells were categorized to GFP, High GFP and Low GFP expression cells. 100000 events were recorded for GFP-positive singlets. High-GFP cells were the top 12% of cells in the GFP-expressing cells and Low-GFP cells were the bottom 12% of

cells in the GFP-expressing cells. 10000 cells from each of the High GFP and Low GFP groups were sorted for recovery.

2.13 Natural product extraction and characterization

2.13.1 Violacein

- Violacein extraction

The violacein-expressing yeast strains were cultured in a suitable selection medium for 5 days at 30°C. 3 ml culture was transferred to a 15 ml Falcon tube, spun down at 5000 g for 5 min and washed twice with ddH_2O. The samples were boiled with 1 ml methanol at 90°C for 10 min and then spun down at top speed for 20 min to 30 min to get rid of the cell debris. The supernatant was transferred to a glass HPLC vial through a 0.2 µm filter. The final volume was standardized to 500 µl with methanol.

- Chemical analysis of violacein

HPLC analysis was performed on an Agilent 1100 series HPLC system, consisting of a quaternary pump (G1311A), a degasser (G1322A), a FLD detector (G1321A) and a DAD detector (G); the column was an Agilent ZORBAX SB-C18, 4.6 x 150 mm, 3.5 µm particle diameter size, 80 Å pore size. Analyses were performed using a gradient of A: H_2O containing 0.1 % TFA and B: MeCN (HPLC grade) containing 0.1 % TFA with a flow rate of 0.8 ml/min. The gradient used was: 0 - 5 min, 5 % B; 5 – 9 min, 5 – 25 % B; 9 – 29 min, 25 – 45 % B; 29 – 40 min, 45 – 100 % B; 40 – 45 min, 100 % B; 45 – 50 min, 100 – 5 % B; 50 – 55 min, 5 % B. Retention times (RT) were denoted in min. Absorption at 210 nm, 220 nm, 256 nm, and 575 nm were recorded. A full absorption spectrum, from 230 nm to 700 nm, was recorded where peaks were observed in the chromatogram.

- Violacein calibration curve

Violacein was purchased from Sigma-Aldrich (V9389-1MG). Solutions with various concentrations (25, 12.5, 6.25, 3.13, 1.56, 0.78, 0.39 and 0.20 mg/l) of commercial violacein were prepared in methanol and analyzed by HPLC. The retention time of violacein was determined to be 21.9 min and the area under the peak was calculated to generate the calibration curve.

2.13.2 Carotenoids

- Carotenoid extraction

The yeast strains were cultured in 200 ml flasks at 30 °C, 220 r.p.m, for 48 h. The OD_{600} of all the yeast strains was recorded and 10 ml of culture was collected by centrifugation at 15000 g for 1 min and washed twice with deionized water. The cells were broken with glass beads (1 g 0.50 to 0.75 mm) by vortexing for 3 min, then 600 µl sorbitol (1 M) and 25 µl lyticase (25 mg/ml) were added. The mixture was vortexed for 10 min, followed by 1 h incubation at 37 °C at 220 r.p.m. 1.25 ml 0.2% (wt/vol) pyrogallol dissolved in methanol was added and mixed by vortexing for 10s. 700 µl 60% (wt/vol) KOH was added for saponification and the mixture was vortexed for 10 s, followed by 1 h incubation at 75°C with vortexing every 15 min. The carotenoid fraction was extracted with 5-10 ml hexane by mixing and left at room temperature overnight with the vertical mixing apparatus. The extraction mixture was centrifuged at 3000 g for 5 min. The hexane layer was transferred to new microcentrifuge tubes and vacuum-dried for 30 min at 60 °C. The cell pellets were re-extracted with hexane extraction. The vacuum-dried extracts were re-dissolved in isopropanol and the whole mixture was cleared by ultrasonication. The cell debris was spun down and the supernant was transferred to a new tube followed by lyophilization. The sample pellet was finally resuspended in 100µl isopropanol and 2 µl of sample was injected to LC-MS equipment for chemical analysis.

- Chemical analysis of lycopene and β-carotene

Separation of lycopene and β-carotene and chemical analysis were performed by the ACQUITY UPLC I-Class system (Waters) and the QTRAP 4500LC-MS/MS system (SCIEX). The column for lycopene and β-carotene separation was an ACQUITY UPLC BEH C18 Column, 130Å, 1.7 µm, 2.1 mm X 50 mm. Analyses were performed using a gradient of A: 100% methanol and B: 50% methanol and 50% isopropanol with a flow rate of 0.3 ml/min. The gradient used was: 0 – 0.5 min, 30 % B; 0.5 – 1 min, 98 % B; 1 – 4 min, 98 % B; 4-4.1 min, 30 % B; 4.1 – 5 min, 30 % B. Retention times (RT) are denoted in min. The parameters for MS were as follows. λ source/gas parameters: ion spray

voltage: 5500; curtain gas: 35, nebulizer gas (gas 1) 55, ion Source GS2 (gas 2) 55; interface heater temperature (IHT): 550. λ compound parameters: de-clustering potential: 30; entrance potential: 10; collision cell exit potential (CXP): 10.

- Lycopene and β-carotene calibration curve

Lycopene and β-carotene were purchased from Sigma-Aldrich. Solutions with various concentrations (2.50, 1.00, 0.75, 0.50, 0.20, 0.10 and 0.05 mg/L of commercial lycopene and 1.00, 0.75, 0.50, 0.20, 0.10, 0.05 and 0.01 mg/L of commercial β-carotene) were prepared in isopropanol and analyzed by LCMS. The retention time of lycopene was determined to be 1.5 min and that of β-carotene was determined to be 1.7 min. The area under the peak was calculated to generate the calibration curve.

2.14 ImageJ assisted cell-counting

- Photo preparation

The ImageJ software was used to facilitate the counting of white and red colonies after transformation. Plates were put at 4°C to develop red color. Each transformation plate was imaged by a Canon digital SLR camera and a light box. For picturing the excision rate test plates, a white background was used for imaging red colonies; a dark pink background was used for imaging white colonies. For picturing the integration rate test plates, a white background was used for red colonies, and a black background was used for white colonies.

- Photo editing and cell number counting

Each plate image was loaded into ImageJ. The “Oval” selection tool was used to select the extent of the plate and the “clear outside” function in the edit menu was used to crop out the background and the edge of the plate. For counting the number of red colonies in a white background, the “Subtract Background” function in the process menu was performed; “light background” and “separate color” with a rolling bar radius of 10-30 pixels was selected and adjusted depending on picture quality. For counting the number of white colonies on a dark background, “Invert” in the edit menu was first used to change it to a light background and “Brightness/Contrast” was adjusted to lower the noise first and then “Subtract Background” was executed for

the red colonies. The bit depth of the image was changed to 16-bit and the "Threshold" was adjusted to make clear black spots on a white background. The "Watershed" function in the Binary menu was performed to detach any merged colonies. Any additional noise was removed using the paintbrush tool. Finally, the "Analyze Particles" function was used in the analyze menu to calculate the colony number. The size option and shape option were further adjusted to get rid of any non-colony noises in the graph. "Show Outline" was chosen to monitor the counting of the colonies.

2.15 Antibiotic function test assays

2.15.1 Overlay assay

- Yeast culture and overlay culture preparation

Yeast cells were grown on agar plates for 2-4 days at 30°C to form colonies before pouring the overlay. The density of colonies was limited to 100-200 colonies/plate so that the buffering capacity of the top agar was not overwhelmed. A fresh overnight (16-24 h) culture of *Bacillus subtilis* 168 was prepared at 30 °C in LB medium ready for the overlay day.

- *B. subtilis* overlay assay

The overlay medium was heated until just bubbling in a microwave and 7-8 ml of the medium was dispensed into sterile test tubes and placed into a pre-warmed 48°C water bath. The overlay medium was cooled to 48°C for more than 15 min to prevent killing the test organism. The plates with yeast growing were pre-warmed in the incubator and taken out when the overlay medium was ready. The tubes with overlay medium were removed one at a time from the water bath to avoid cooling too fast which can cause uneven overlays. Fresh overnight *B. subtilis* culture was diluted 1:200 into the overlay medium, mixed gently by vortexing and poured onto the yeast agar plate. The yeast plate with the overlay medium was left at room temperature for more than 15 min to allow the overlay to cool and harden. The plate/overlay was incubated upside-down at 30°C for 16-24 h and photographed after the incubation.

2.15.2 Antibiotic susceptibility disk diffusion assay

Before use, MH plates were pre-warmed at 37°C for 20 min. 50 µl of an overnight culture of *B. subtilis* was spread on the MH plates using glass beads and left to dry. The plate was quartered using a marker and the center of each quarter was labelled with the antibiotics to be put on that location. Under sterile conditions (with the Bunsen burner) and using tweezers, the appropriate volume of antibiotics was added to the sterile disks and then placed at the center of the quarter on the plate (antibiotics faced down). The plate was incubated in 30°C for overnight culture and photographed.

Chapter 3 Recombinase SCRaMbLE toolkit for the synthetic yeast genome

3.1 Overview of SCRaMbLE toolkit and circuit design

In the design of the synthetic yeast chromosomes, the Cre/*loxP*$_{sym}$ recombination system was used for the purpose of SCRaMbLE (Dymond et al., 2011). Random genome variations can be generated by the SCRaMbLE system and directed evolution can be further achieved by applying different selection pressure. One main purpose of performing SCRaMbLE on the synthetic chromosomes is to achieve shorter, more succinct and even minimal genomes for *S. cerevisiae*. Another major design strategy for the synthetic yeast genome is to transfer and isolate the unstable elements – tRNA genes – from the local chromosome locations and build a tRNA Neochromosome so that the cellular roles of tRNA genes can be studied specifically (Figure 3.1).

There are 275 tRNA genes grouped into 42 families dispersed in the yeast nuclear genome. In theory, only 64 tRNAs with corresponding anti-codons are needed for translation; and even less tRNA genes are needed if tRNA wobble is taken into account. Thus, screening for essential tRNA genes can help to reduce the total length of the chromosomes. Apart from transferring amino acids, tRNA genes also seem to play other roles in the genome. For instance, tRNA genes have been proved to work as an insulator in blocking the spread of silent chromatin (Dhillon et al., 2009; Donze & Kamakaka, 2001; Lynch & Rusche, 2010) and to participate in regulation of gene transcription (Donze & Kamakaka, 2001). Retrotransposons like Ty1 also target the upstream regions of tRNA genes (Mularoni et al., 2012), which may lead to transposition of tRNA genes. To study the relationship between tRNA genes and their local environments and additional functions in the genome, it would be interesting to transfer them all to a Neochromosome and study the effects on yeast phenotypes. Furthermore, tRNA genes are fragile genome sites that can trigger DNA breakpoints (Di Rienzi, Collingwood, Raghuraman, & Brewer, 2009). RNA polymerase III and DNA

polymerase may meet and collide and lead to single-strand breaks if the tRNA is transcribed heavily during DNA replication. Moreover, there are only 17 tRNA genes on average in each chromosome in the natural yeast genome; so we can see whether there will be any difference if all of the tRNA genes are put together in a single chromosome and whether the tRNA chromosome can be further evolved by SCRaMbLE. There could be problems of chromosome stability of the tRNA Neochromosome and SCRaMbLE could be used to select relatively more stable elements and improve the fitness.

For the purpose of specific SCRaMbLE of the tRNA Neochromosome, an orthogonal recombination system to Cre/loxP needs to be developed and fine-tuned for expression and function. In this chapter, the design, construction and hormone regulation of five tyrosine recombinases are introduced, followed by SCRaMbLE tests on synthetic chromosomes and a partially synthesized tRNA array. With the success of the orthogonal recombinase device, the activation of SCRaMbLE of either normal synthetic chromosomes or the tRNA Neochromosome can be precisely controlled.

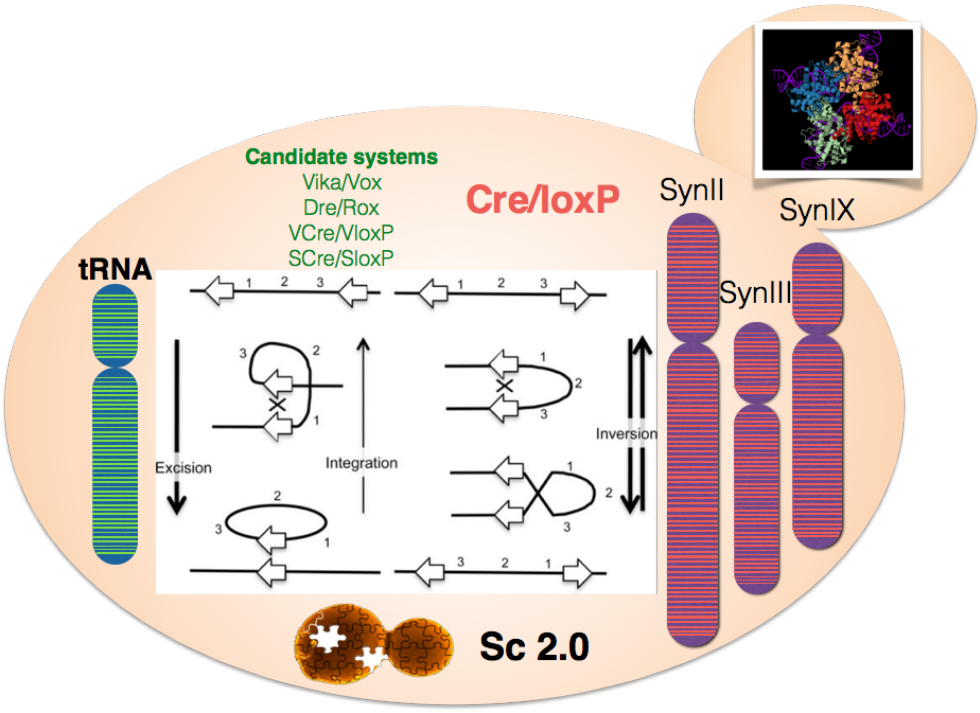


Figure 3.1| Overview scheme for building orthogonal SCRaMbLE systems for tRNA Neochromosome and normal synthetic chromosomes. Cre/*loxP* was used for SCRaMbLE of normal synthetic chromosomes and four more new recombination systems were chosen as candidates for SCRaMbLE of the tRNA Neochromosome.

3.1.1 Choices of tyrosine recombinases

Many new recombinases have been discovered by genome mining in recently years, including Dre/rox (Anastassiadis et al., 2009), Vika/Vox (Karimova et al., 2013), VCre/VloxP (Suzuki & Nakayama, 2011), SCre/SloxP (Suzuki & Nakayama, 2011), etc. There are many conserved domains among the recombinases while still recognizing unique target sequences (Figure 3.2 and Table 3.1). In this study, these four new recombinases together with the well-studied Cre recombinase were chosen as candidate systems for the SCRaMbLE of synthetic chromosomes. Since the recombinases were originally from either bacteriophage or *Vibrio* species, codon optimization is necessary for customized efficient expression of these recombinases in yeast. Cre was amplified by PCR from a CreEBD expression vector from a previous study (Lindstrom & Gottschling, 2009); Dre was identified from phage D6 and the coding sequence was obtained from DDBJ/EMBL/GenBank database with the accession number AY753669 (Sauer & McDermott, 2004). VCre was identified from *Vibrio sp*. 0908 and the sequence was obtained from GenBank with the accession number ABX77110.1. Vika originates from a degenerate bacteriophage of *Vibrio coralliilyticus* ATCC BAA-450 and the sequence was obtained from GenBank with the protein ID WP_006958771.1. SCre was identified from *Shewanella sp.* ANA-3 and the sequence was obtained from GenBank with the accession number ABK50591.1. After obtaining the original sequence of the recombinases, the sequences were codon-optimized for *S. cerevisiae* by the online software GeneDesign (Richardson, Wheelan, Yarrington, & Boeke, 2006). Finally, Dre was synthesized by the "Build-A-Genome" protocol described in chapter 2.3.1.1. VCre, Vika and SCre were ordered as GeneArt Strings from Thermo Fisher Scientific.

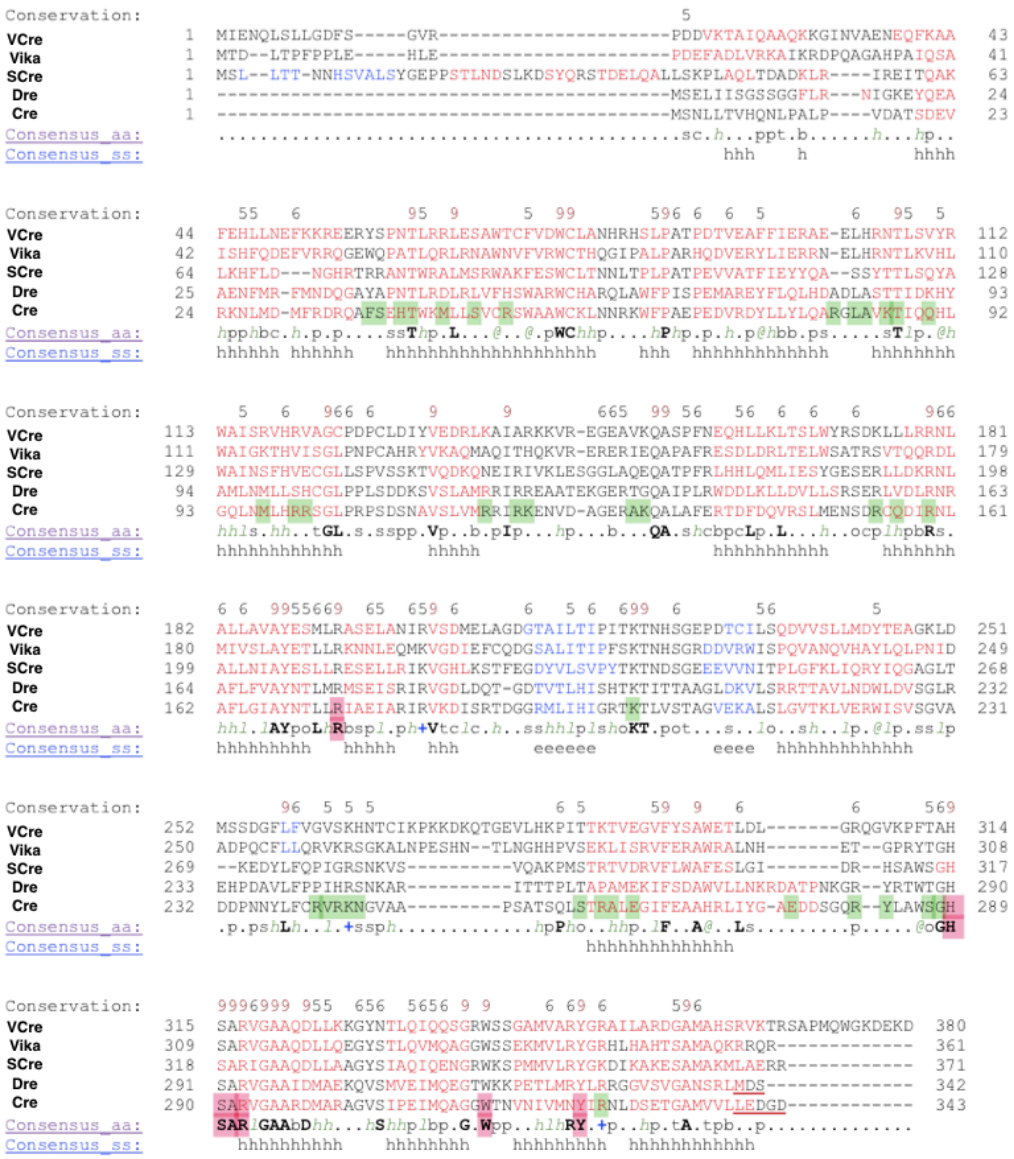


Figure 3.2|Amino acid sequence comparison of the candidate recombinases. The alignment was done by PROMALS3D online server, with consensus amino acid and secondary structure marked as above.

Table 3.1| Recombination sites of candidate recombination systems

Name	Target site	Length	Site sequence
Cre	loxP	34 bp	ATAACTTCGTATAGCATACATTATACGAAGTTAT
Dre	rox	32 bp	TAACTTTAAATAATTGGCATTATTTAAAGTTA
VCre	VloxP	34 bp	TCAATTTCCGAGAATGACAGTTCTCAGAAATTGA
SCre	SloxP	34 bp	CTCGTGTCCGATAACTGTAATTATCGGACATGAT
Vika	Vox	34 bp	AATAGGTCTGAGAACGCCCATTCTCAGACGTATT

3.1.2 Transcriptional regulation of gene expression

After the sequences of the candidate recombinases were decided, the next step was to achieve tunable and controllable expression in *S. cerevisiae.* As a first step to test whether the recombinases could actually drive recombination between target recombination sites (RS), strong expression was necessary; for SCRaMbLE application, fine-tuned moderate expression is necessary since excessively high expression of SSRs will lead to high loss rate of essential genes and expression at too low a level will not be enough to generate sufficient genome variation in a reasonable time period. For transcriptional regulation, different promoter and terminator combinations were used for recombinase transcription unit (TU) construction. A strong constitutive promoter P_{TDH3}, a strong inducible promoter P_{Gal1} and a weak cell cycle-related promoter P_{SCW11} were chosen for transcription initiation. A constitutive terminator of Gal1 and a terminator of MFA2 that can reduce the half-life of mRNA (Muhlrad & Parker, 1992; Yamanishi et al., 2013) were chosen for transcription termination and controlling mRNA stability (Figure 3.3a). A yeast centromere plasmid with *LEU2* marker was chosen as the backbone for the SSR expression device. The details of the constructs were listed in Table 3.2.

Table 3.2 Recombinase expression devices

Strain ID	Plasmid ID	Description
LWe044	pWL001	P_{Gal1}-Dre-T_{Gal1}-pRS415
LWe045	pWL002	P_{Gal1}-Dre-T_{MFA2}- pRS415
LWe046	pWL003	P_{TDH3}-Dre- T_{Gal1}- pRS415
LWe047	pWL004	P_{TDH3}-Dre- T_{MFA2}- pRS415
LWe069	pWL010	P_{Gal1}-DreEBD- T_{Gal1}- pRS415
LWe070	pWL011	P_{Gal1}-DreEBD- T_{MFA2}- pRS415
LWe071	pWL012	P_{TDH3}-DreEBD- T_{Gal1}- pRS415
LWe072	pWL013	P_{TDH3}-DreEBD- T_{MFA2}- pRS415
LWe073	pWL014	P_{SCW11}-Dre- T_{MFA2}- pRS415
LWe074	pWL015	P_{SCW11}-DreEBD- T_{MFA2}- pRS415

Strain ID	Plasmid ID	Description
LWe094	pWL024	P_{SCW11}-Vika- T_{MFA2}- pRS415
LWe095	pWL025	P_{SCW11}-Vcre- T_{MFA2}- pRS415
LWe096	pWL026	P_{SCW11}-Scre- T_{MFA2}- pRS415
LWe109	pWL033	P_{Gal1}-VCre- T_{MFA2}-pRS415
LWe110	pWL034	P_{TDH3}-VCre- T_{MFA2}- pRS415
LWe111	pWL035	P_{TDH3}-VCreEBD- T_{MFA2}-pRS415
LWe112	pWL036	P_{SCW11}-VikaEBD- T_{MFA2}-pRS415
LWe113	pWL037	P_{SCW11}-SCreEBD- T_{MFA2}-pRS415
LWe140	pWL040	P_{Gal1}-Vika- T_{MFA2}-pRS415
LWe141	pWL041	P_{Gal1}-VCre- T_{MFA2}-pRS415
LWe142	pWL042	P_{Gal1}-SCre- T_{MFA2}-pRS415
LWe143	pWL043	P_{SCW11}-Dre(+GD)EBD- T_{MFA2}-pRS415
LWe160	pWL044	P_{SCW11}-DrePBD- T_{MFA2}-pRS415
LWe199	pWL061	P_{SCW11}-VikaEBD- T_{MFA2}-pRS415
LWe200	pWL062	P_{SCW11}-VCreEBD- T_{MFA2}-pRS415
LWe201	pWL063	P_{SCW11}-SCreEBD- T_{MFA2}-pRS415
LWe202	pWL064	P_{SCW11}-VCre- T_{MFA2}-pRS415
LWe217	pWL078	P_{TDH3}-CrePBD- T_{MFA2}-pRS415
LWe218	pWL079	P_{TDH3}-DrePBD- T_{MFA2}-pRS415
LWe219	pWL080	P_{TDH3}-VikaPBD- T_{MFA2}-pRS415
LWe220	pWL081	P_{TDH3}-SCrePBD- T_{MFA2}-pRS415

3.1.3 Post-translational hormone-controlled regulation of gene function

Once the basic function of a recombinase in yeast was confirmed, a post-translational regulation system was applied for activation and deactivation of the recombinase. As is introduced in section 1.3.1.4, post-translational regulation is based on the application of nuclear receptors that can sense small hormone molecules via their ligand binding domain (LBD). In this study, functional recombinase was fused with either an estradiol-binding domain (EBD) or a progesterone-binding domain (PBD) for

functional regulation. When the hormones are absent, the ligand-binding domain is bound with heat-shock proteins such as Hsp90 in the cytoplasm, and therefore the fused recombinase is also restricted to the cytoplasm and prohibited from acting on DNA in the nucleus; the introduction of small hormone molecules will cause conformational change and interrupt the binding between LBD and heat shock protein, resulting in the release of the recombinase fusion protein from the cytoplasm (Figure 3.3b). With the help of a nuclear localization sequence (NLS), the recombinase is able to enter the nucleus and perform recombination on target DNA.

3.1.4 Function testing devices and workflow

To test the basic function of the recombinases and the effect of regulation, a double recombination site flanking *URA3* reporter device was designed. The reporter was constructed with a centromere plasmid backbone with *HIS3* (pRS413) as auxotrophic marker (Figure 3.4). Five reporter devices, for each recombination system, were constructed: pRS413-loxP-URA3-loxP, pRS413-rox-URA3-rox, pRS413-VloxP-URA3-VloxP, pRS413-SloxP-URA3-SloxP and pRS413-Vox-URA3-Vox (Table 3.3). Recombination between the sites flanking *URA3* will result in loss of the *URA3* expression unit. To measure and evaluate how efficient the recombination process is, auxotrophic selection and spotting assay were applied to indicate the degree of loss of the *URA3* unit. The workflow is illustrated in figure 3.5. The reporter device was first transformed to BY4741 and selected on SC-His plates; then the recombinase expression device was further transformed into this strain and selected on SC-His-Leu plates. A primary replica-plating assay on SC-His-Leu-Ura triple dropout media was done for qualitative function tests. By comparing the growth status of a colony on both SC-His-Leu and SC-His-Leu-Ura plates, a rough degree of recombination extent can be estimated. For the colonies that grow on SC-His-Leu but not on SC-His-Leu-Ura, plasmid recovery, digestion verification and DNA sequencing were performed to finally confirm the loss of the *URA3* unit and whether recombination has taken place between the recombination sites. For quantitative recombination rate measurement, several transformants were inoculated into SC-His-Leu liquid

medium and spotted on SC-His-Leu and SC-His-Leu-Ura for survival rate comparison. For hormone-induced recombination, the induction could be either on plates or in the liquid medium. Plate induction starts immediately when the cells are plated and induction will continue during the growth of the colonies, while by liquid induction the period of activation can be precisely controlled. The plate induction and liquid induction methods were described in chapter 2.5.

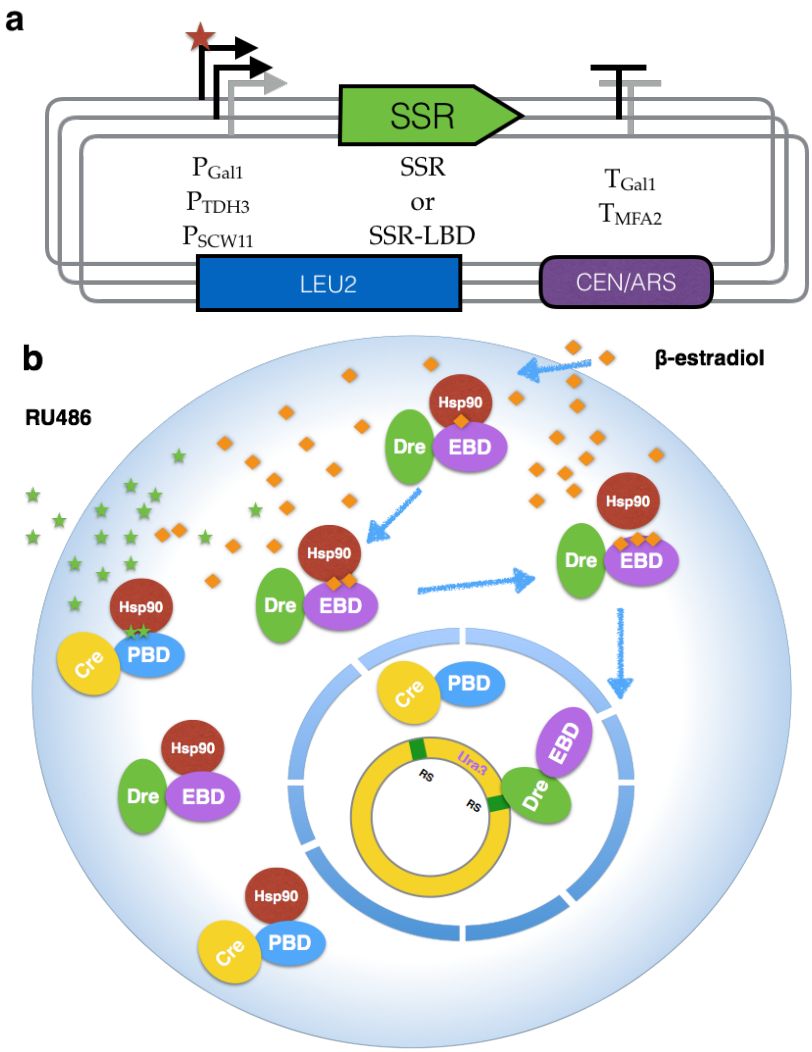


Figure 3.3 | Tuning and control of recombinase function. a. Transcriptional regulation of recombinases. Constitutive and inducible, strong and weak promoters, normal and mRNA stability-related terminators were used for tuning SSR transcriptional rate. **b.** Post-translational regulation of recombinases. Different hormone-binding domains from nuclear receptors were fused with SSRs for post-translational spatial control of recombinase function. Hsp90: heat shock protein 90. EBD: estrogen-binding domain. PBD: progesterone-binding domain. RS: recombination site.

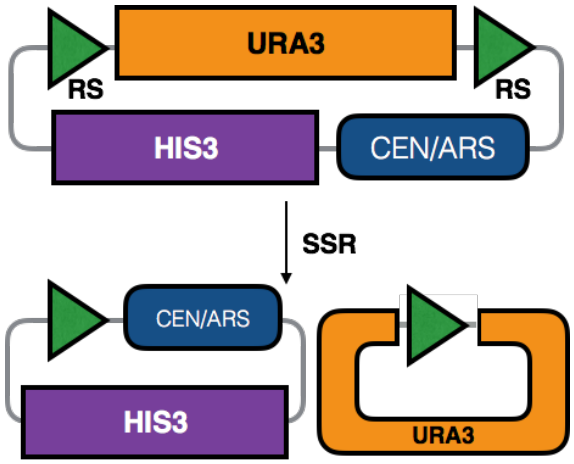


Figure 3.4|Design of recombinase function reporter device. A *URA3* cassette flanked by recombination sites is assembled in pRS413. When recombination happens between the two recombination sites, the *URA3* expression unit will be lost. Uracil drop-out medium is used to indicate the presence of the *URA3* expression unit.

Table 3.3 Function reporter devices

Strain ID	Plasmid ID	Description
LWe051	pWL008	roxURA3rox-pRS413
LWe090	pWL021	VoxURA3Vox-pRS413
LWe091	pWL022	VloxPURA3VloxP-pRS413
LWe092	pWL023	SloxPURA3SloxP-pRS413
LWe224	pWL085	loxPURA3loxP-pRS413

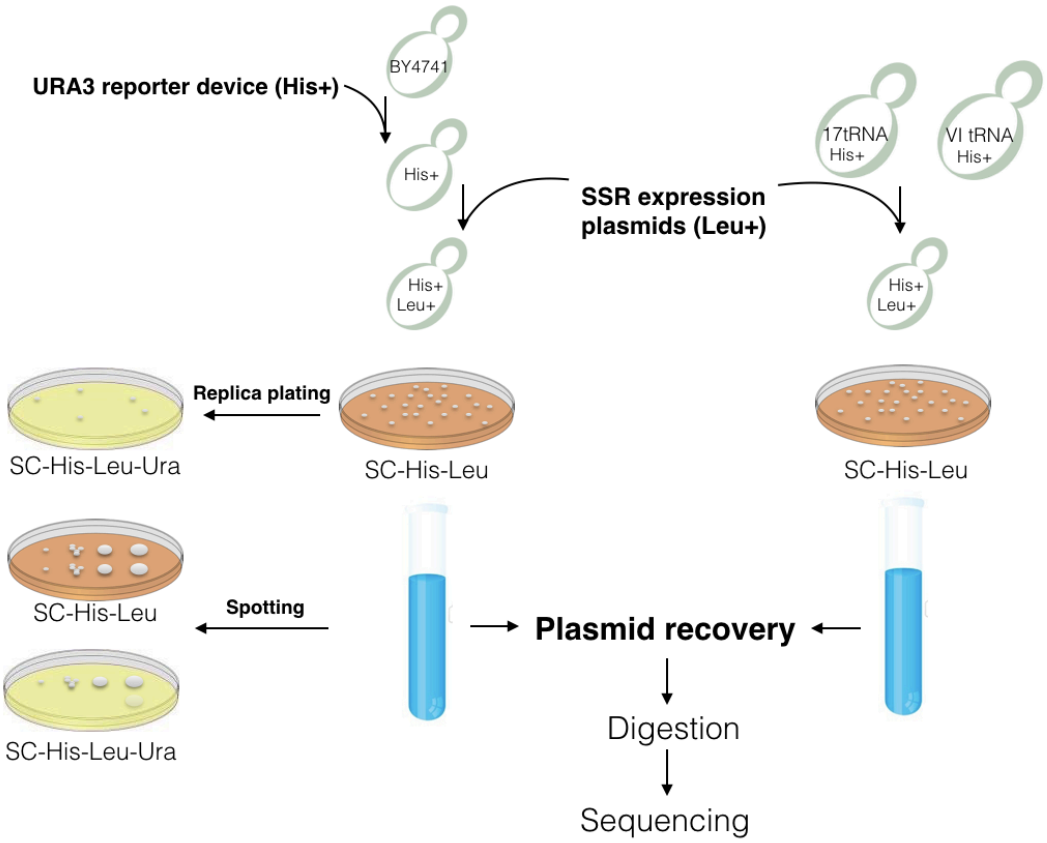


Figure 3.5|Recombinase function testing workflow. Double transformants of *URA3* reporter device and Dre expression device are first selected on a SC-His-Leu plate. Then both replica plating assay and spotting assay are performed to compare the colony forming status on SC-His-Leu and SC-His-Leu-Ura plates. After primary phenotype comparison, subsequent genotype verification is performed to verify that recombination indeed happens between the recombination sites. Plasmid recovery, DNA digestion assay and sequencing are performed for recombination confirmation.

3.2 Results

3.2.1 Functional test of Dre recombinase by phenotype verification

As described above, a *URA3* reporter device was designed for basic function test of the recombinase. In the case of Dre recombinase, rox flanking device pWL008 was first transformed to BY4741, followed by further transformation with pRS415 or a Dre expression device like pWL003 and pWL004. After transformation selection on SC-His-Leu plates, the plates were replica-plated to SC-His-Leu-Ura plates. As shown in figure 3.6, the negative control strain without *URA3* selection marker did not grow on SC-His-Leu-Ura plates; the positive control strain with empty vector pRS415 grow

well on the triple dropout plate, while test strains with P_{TDH3}-Dre expression devices have dramatic reduction in growth on the triple dropout plate, indicating a high possibility of *URA3* cassette loss in the Dre expressing strains.

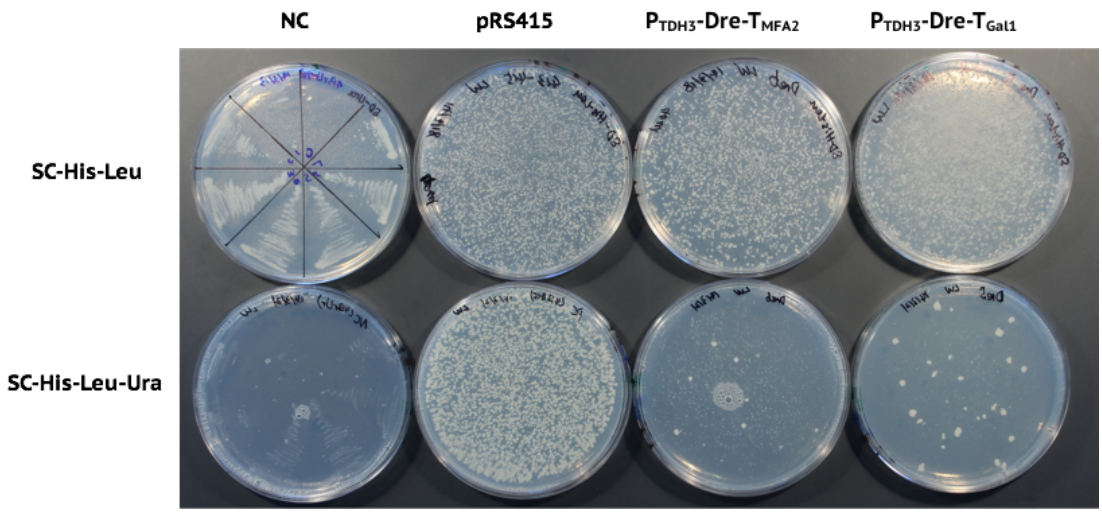
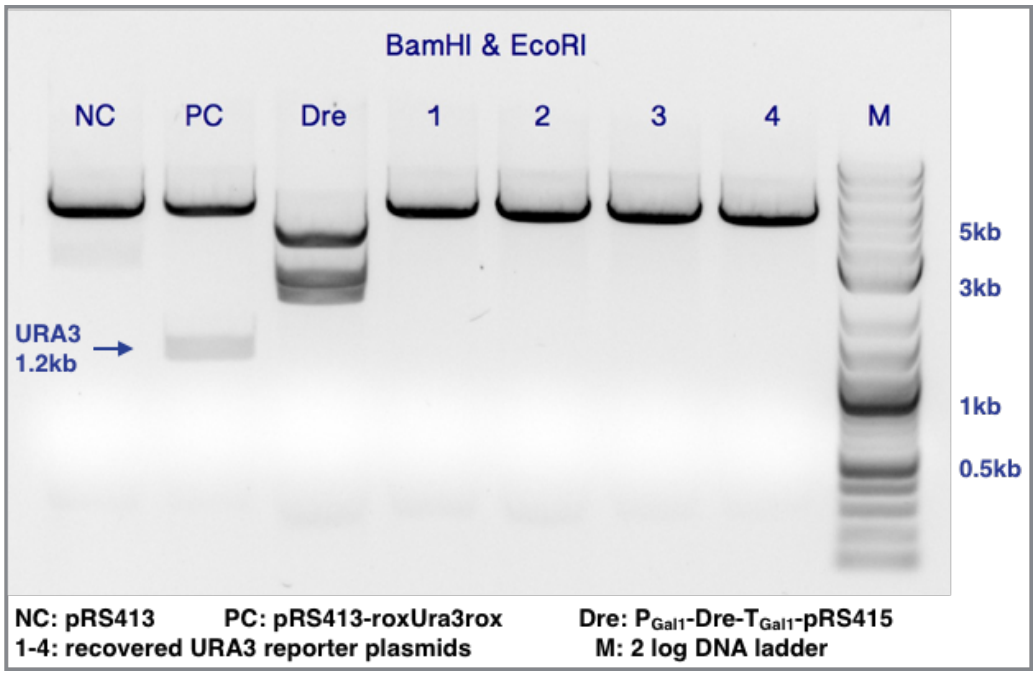


Figure 3.6| Replica plating assay for Dre function test. NC: Negative control strain with *HIS3* and *LEU2* marker plasmids, without *URA3* marker. Reporter device with empty vector pRS415 grows well on SC-His-Leu-Ura plate, while most colonies of strains with Dre expression devices cannot survive on SC-His-Leu-Ura plates.

To further genotypically confirm whether the phenomenon is caused by the loss of *URA3* and whether the recombination indeed happened between the rox sites, four single colonies that grew on SC-His-Leu but not on SC-His-Leu-Ura were chosen for genotype verification. The plasmids in the four colonies were recovered by yeast plasmid extraction, re-transformed to *E. coli* DH5α, mini-prepped and digested with BamHI and EcoRI. In the digestion assay, three controls were used for comparison: a pRS413 empty vector, the pRS413-rox-URA3-rox and Dre expression device (Figure 3.7a). BamHI and EcoRI digestion of the original reporter device should result in a 6kb backbone band and a 1.2kb URA3 band. From the digestion result, we can see that, the digestion pattern of the four recovered plasmids were the same as pRS413 but not the *URA3* reporter device that has the 1.2kb band, indicating the loss of the *URA3* cassette. Dre expression device was also used as control to rule out any recovered Dre expression plasmid. As a final confirmation, the four recovered plasmids were sent for sequencing and the result shows that there is only one rox

site left on the plasmid, which is expected as the result of recombination between the rox sites (Figure 3.7b).

a.



b.

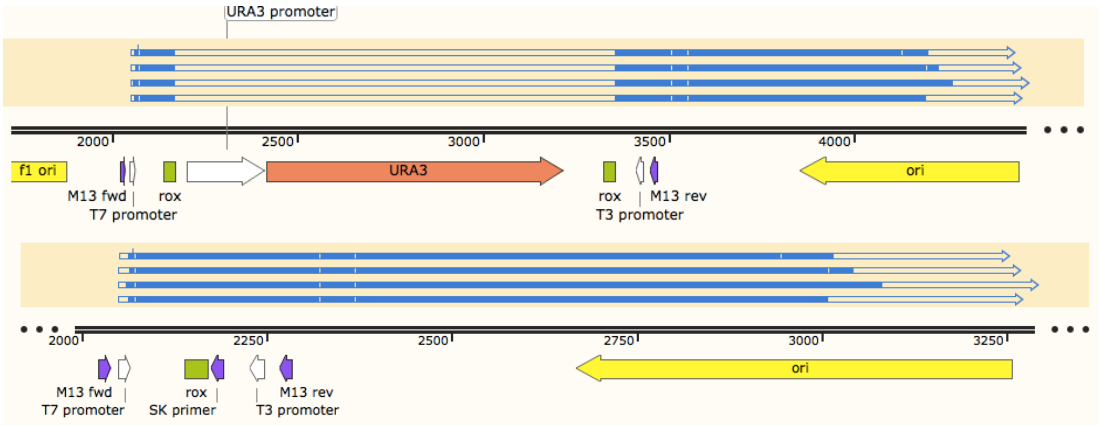


Figure 3.7| Recombination confirmation of the testing device. a. Digestion pattern of recovered recombination plasmid. Digestion of pRS413-roxURA3rox plasmid by BamHI and EcoRI will separate the *URA3* cassette from the pRS413 backbone. If *URA3* cassette exists, there should be a 1.2kb band and a 6kb band. The digestion result shows that the four recovered plasmids have the same digestion pattern as pRS413 without the 1.2kb band, indicating the loss of *URA3* cassette for all the recovered plasmids. The Dre expression device was also used as control to rule out recovered Dre expression plasmid. **b.** Sequencing result of recovered reporter plasmids. The result shows that only one rox site was left on the plasmid, proving that the recombination between rox sites was successfully activated by Dre recombinase.

3.2.2 Evaluation of recombination efficiency

Through the above assays, the recombination between rox sites by the function of Dre was confirmed by sequencing. To further quantify the recombination efficiency in a population, serial dilution and spotting assay was performed on strains with different Dre-expressing devices. Strains with P_{TDH3}-Dre and P_{SCW11}-Dre devices were cultured overnight in SC-His-Leu glucose liquid culture and those with P_{Gal1}-Dre devices were cultured in SC-His-Leu galactose. They were then spotted on SC-His-Leu and SC-His-Leu-Ura plates with glucose. The control strain with the empty vector pRS415 grew well on both plates while those with P_{TDH3}-Dre, P_{Gal1}-Dre or P_{SCW11}-Dre devices all failed to grow on SC-His-Leu-Ura plates (Figure 3.8).

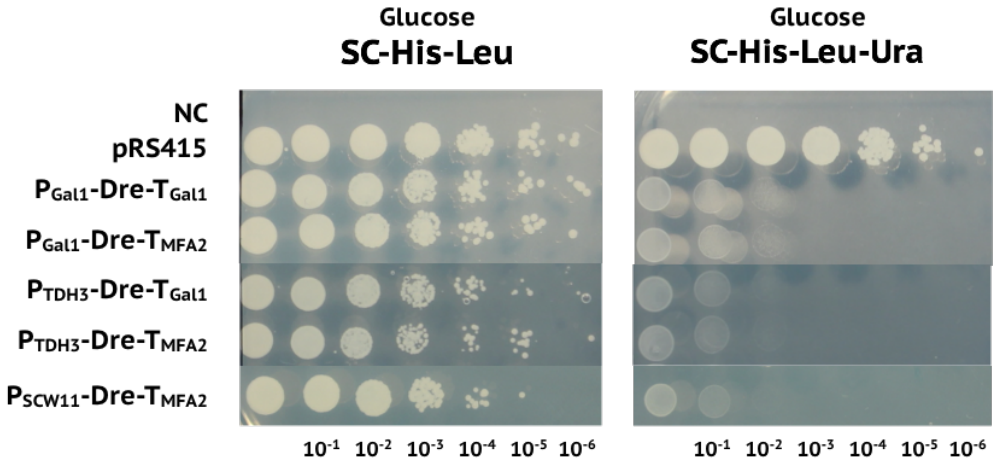


Figure 3.8| Spotting assay of Dre expression devices. Five Dre expression devices were used for function testing, including P_{Gal1}-Dre with either T_{Gal1} or T_{MFA2}, P_{TDH3}-Dre with either T_{Gal1} or T_{MFA2} and P_{SCW11}-Dre with T_{MFA2}. Strains with P_{Gal1}-Dre were cultured in galactose liquid medium and the rest were cultured in glucose liquid medium overnight. The spotting result shows that strains with reporter devices and pRS415 empty vector grows well on both SC-His-Leu and SC-His-Leu-Ura plates; strains with all Dre expression devices grew well on SC-His-Leu devices but did not form colonies on the triple dropout plate, indicating super high recombination efficiency of Dre.

This indicates that the recombination activated by Dre is very efficient; even the use of the weak cell-cycle related promoter P_{SCW11} and the T_{MFA2} that should shorten the

half-life of the mRNA did not restrict the function of Dre. This is possibly because although the level of expression of Dre is diversified, the timing of its function is not controlled because P_{TDH3} and P_{SCW11} are constitutive promoters but not inducible. Though P_{Gal1} is inducible by galactose, the control of P_{GAL1} is very leaky-even without galactose induction the P_{Gal1}-Dre device can still generate enough recombinase to excise the *URA3* cassette (data not shown). Therefore, a more precise control of Dre function is necessary.

3.2.3 Orthogonality test for Dre and Cre

Before introducing the hormone-induction of recombinase function, another fundamental test for SCRaMbLE of the synthetic chromosome and the Neochromosome is the orthogonality test for the two recombinases. The function of Cre was already demonstrated in a previous study (Lindstrom & Gottschling, 2009) and the basic function of Dre was demonstrated in this study. Once the orthogonality test of the two recombinases is passed, the two recombination systems can be applied to SCRaMbLE the synthetic chromosomes. In this test, the rox-URA3-rox and loxP-URA3-loxP reporter plasmids were first transformed into BY4741 strain; then a Dre expressing device and a Cre-expressing device were transformed into both rox-URA3 and loxP-URA3 reporter containing strains. The spotting assay shows that the strain with roxURA3rox device and Dre did not survive on the SC-His-Leu-Ura plate, while that with Cre grew well on the triple dropout plate, indicating that Dre can work on rox sites but Cre cannot; the strain with loxPURA3loxP device and Cre did not grow on the triple dropout plate but that with Dre survived, indicating that Cre can work on loxP site but Dre cannot. This assay proves that there is no cross-talk between the Dre/rox and Cre/loxP systems and the two orthogonal systems can be applied for selective SCRaMbLE of normal chromosomes and the tRNA Neochromosome.

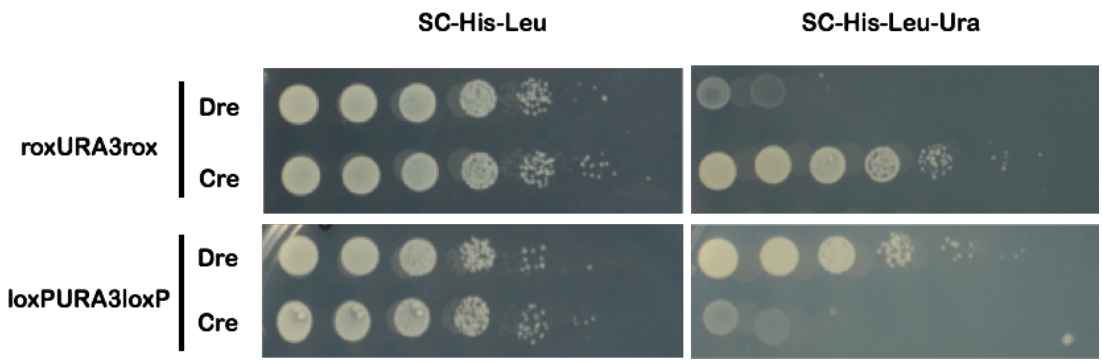


Figure 3.9| Orthogonality test of Dre and Cre function by spotting assay. Two reporter devices were used for the assay. For strains with rox-URA3 reporter device, those with Dre failed to grow on SC-His-Leu-Ura plate while those with Cre survived; for strains with loxP-URA3 reporter device, those with Dre survived while those with Cre did not. This proves that only Dre can work on rox sites and Cre can only work on loxP sites. The two systems are orthogonal. Dre: P_{SCW11}-Dre; Cre: P_{SCW11}-CreEBD 1μM β-estradiol induced.

3.2.4 Hormone regulation of Dre and Cre

3.2.4.1 Functional test of DreEBD

After the basic recombination function of Dre was confirmed and its orthogonality with Cre was also tested, it was time to enable hormone regulation of the recombination. Dre was fused to the same EBD which had been demonstrated to work with Cre. A DreEBD expression cassette was constructed by Golden Gate assembly to fuse EBD to the C-terminus of Dre with a 3 aa (LEP) linker in the same acceptor vector pRS415, and the same *URA3* reporter device in pRS413 was applied. The inducibility testing workflow for DreEBD is similar to the basic function test of Dre, including both replica plating assay with plate induction and spotting assay with both plate and liquid induction (Figure 3.10). For the replica plating assay, SC-His-Leu-Ura with and without β-estradiol were used for survival rate comparison. The replica plating assay can primarily indicate whether the EBD can confine Dre in the cytoplasm and release it by adding the hormones; then spotting assay was performed to quantitatively evaluate the induction efficiency. For the spotting assay, single colonies were inoculated into SC-His-Leu liquid medium to be cultured overnight. For plate induction, the cells were re-inoculated to OD_{600} of 1.0 and spotted to three

kinds of plates: SC-His-Leu, SC-His-Leu-Ura and SC-His-Leu-Ura with the corresponding hormone molecule. For liquid induction, the culture was re-inoculated to OD_{600} of 0.3 and the corresponding hormone molecule was added to the liquid medium for induction. After induction for a set time period, the cells were spotted on SC-His-Leu and SC-His-Leu-Ura without hormone. In plate induction, the hormone activation starts right away with the plating and will continue for at least 48h until the yeast colony is formed, while in liquid induction the higher initiating cell density could reduce potential stress on cell growth generated by recombinase expression and the induction time can be more flexibly controlled.

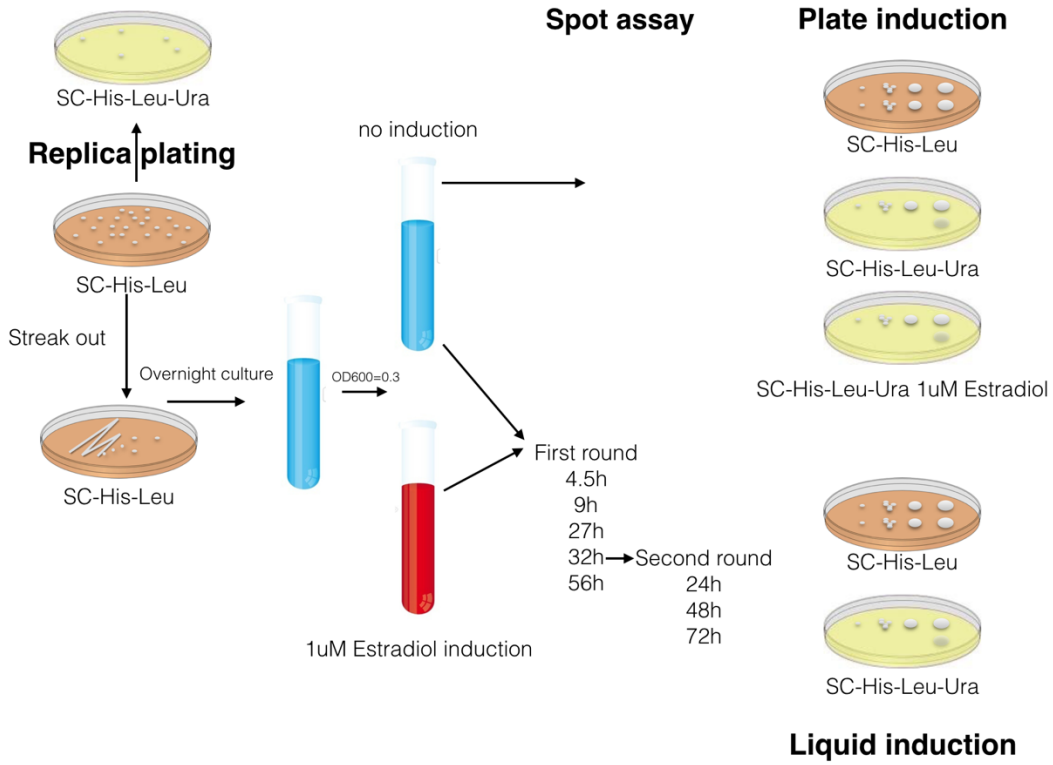


Figure 3.10| Function test workflow for hormone induction of recombinases. The replica plating assay and spot assay were used for hormone induction testing. The former is for primary estimation of inducibility and the latter is for quantitative function test. Replica plating can only be induced on plates while the spot assay can be induced either in liquid or on plate. The time for plate induction has to be at least 48h for colonies to grow up while liquid induction is more flexible and different time points and rounds of induction can be achieved. In the end, colony number after serial dilution can be compared between different strains on different testing plates.

Two DreEBD expression devices, P_{TDH3}-DreEBD and P_{SCW11}-DreEBD, were used for the test, and P_{SCW11}-Dre was used as positive control. The replica plating result is shown in Figure 3.11. On SC-His-Leu plates, all strains grew well; on SC-His-Leu-Ura plate, only strains with P_{SCW11}-Dre have obvious colony reduction while strains with DreEBD fusion protein grew as well as those on SC-His-Leu, indicating no *URA3* loss without β-estradiol induction. On SC-His-Leu-Ura β-estradiol-containing plates, strains with DreEBD fusion protein grew similarly to those on the triple dropout plate with no estradiol, which is not as expected since the estradiol induced strain should have more obvious colony loss compared with those without induction.

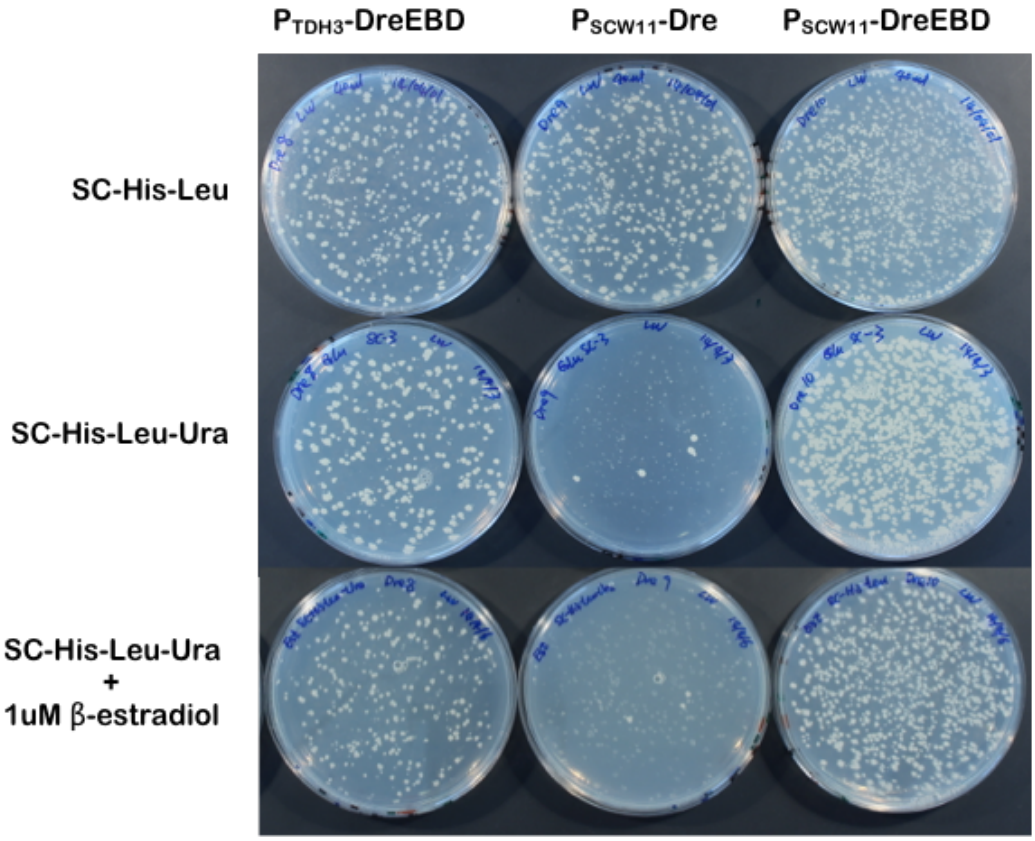


Figure 3.11| Replica plating assay of DreEBD function. All strains grew well on SC-His-Leu plates; the strain with P_{SCW11}-Dre barely grew on the SC-His-Leu-Ura plates indicating the loss of the *URA3* cassette; the strains with P_{TDH3}DreEBD and P_{SCW11}DreEBD grew similarly on SC-His-Leu-Ura plate and SC-His-Leu plate indicating the restriction effect of EBD without induction; while on the β-estradiol induction plate, the strain with DreEBD expressing device still grew similarly to the one without induction, indicating low recombination efficiency.

However, since the replica plating assay can only give a rough comparison before and after induction but cannot quantitatively evaluate the induction efficiency, the result is not yet conclusive to determine whether DreEBD can be induced or not. Therefore, the spotting assay was performed to further evaluate and optimize the inducibility. Following the workflow described above, an 7h liquid induction by 1μM β-estradiol was performed on strains with P_{SCW11}-Dre, P_{SCW11}-DreEBD and P_{TDH3}-DreEBD and spotted to SC-His-Leu plates and SC-His-Leu-Ura plates. The strains with P_{SCW11}-Dre did not grow on triple dropout plate, proving the function of Dre; the colony number of strain with P_{TDH3}-DreEBD and P_{SCW11}-DreEBD was slightly lower on triple dropout plates than on the double dropout: strain with P_{TDH3}-DreEBD has more than 10 colonies on double dropout while it has 6 colonies on triple dropout on level of 10^{-4} dilution; strain with P_{SCW11}-DreEBD have 7 colonies on double dropout plate while have only 1 on triple dropout plate (figure 3.12a). Though the result was not yet conclusive to say whether the induction worked or not, it indicated some clues that the induction might work better with longer induction and more concentrated inducer. This hypothesis was supported by the plate induction result. In contrast to liquid induction, the plate induction is always on during colony forming, which takes more than 24 h, normally 2 to 3 days. We can see that on the 1μM β-estradiol containing triple dropout plate, the colony number of the strain with empty vector is around 10 times that with P_{SCW11}-DreEBD (Figure 3.12b). Though the induction efficiency is not yet as good as CreEBD, it is worth optimizing induction conditions specifically for DreEBD.

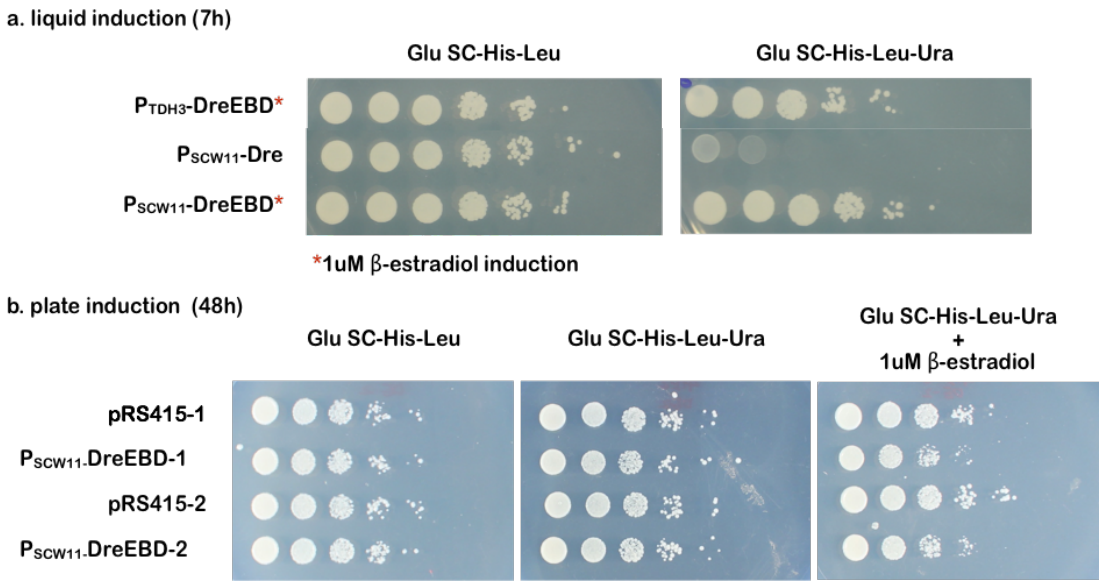


Figure 3.12| Spotting assay of DreEBD function. a. 7h liquid induction result. Strain with P_{SCW11}Dre was used as positive control for Dre function and it worked well since it did not grow on SC-His-Leu-Ura plate; strains with P_{TDH3}DreEBD and P_{SCW11}DreEBD with 1μM β-estradiol have slightly lower colony number on SC-His-Leu-Ura than on SC-His-Leu plate. **b.** 48h plate induction result. Strain with P_{SCW11}DreEBD grew similarly on SC-His-Leu and SC-His-Leu-Ura plate without estradiol, while it has 10-fold colony number reduction on SC-His-Leu-Ura estradiol-containing plate. Empty vector pRS415 was used as negative control and biological replicates were performed for the assay.

3.2.4.2 Induction optimization of DreEBD

The induction condition optimization for DreEBD includes timing, estradiol concentration and linker modification for the fusion protein.

Since obvious colony loss in the 48h plate induction assay was observed, different induction times were tested to compare the induction efficiency. In this assay, P_{SCW11}DreEBD was used. In a first round of liquid induction, samples were taken out and spotted on double dropout and triple dropout plates at 4.5h, 9h, 27h, 32h, 56h. Considering the thermal stability of estradiol and nutritional requirement, all samples were re-inoculated to fresh liquid medium with estradiol at 32h for a second round of induction. The first round induction was continued for another 24h for comparison with a fresh 24h-second round induction. Samples from second round of induction were taken for spotting at 24h, 48h and 72h (Figure 3.10). The result shows that

before 9h, there is no obvious growth difference on double and triple dropout plates; 27h induction shows around 3-fold reduction on triple dropout plate; 32h shows 10-fold reduction (Figure 3.13). Comparing the recombination efficiency of 56h induction in the 1st round and 24h induction in the 2nd round, I found that although the total induction time is the same (56h in total) the recombination rate of 2nd round induction with re-inoculation and fresh β-estradiol is much higher than that of continuous 56h induction (Figure 3.13). Longer induction to 48 h in the 2nd round further increased the recombination rate while 72h induction did not much change the rate compared with 48h induction. It indicates that the effect of β-estradiol decreased with prolonged culture at 37°C.

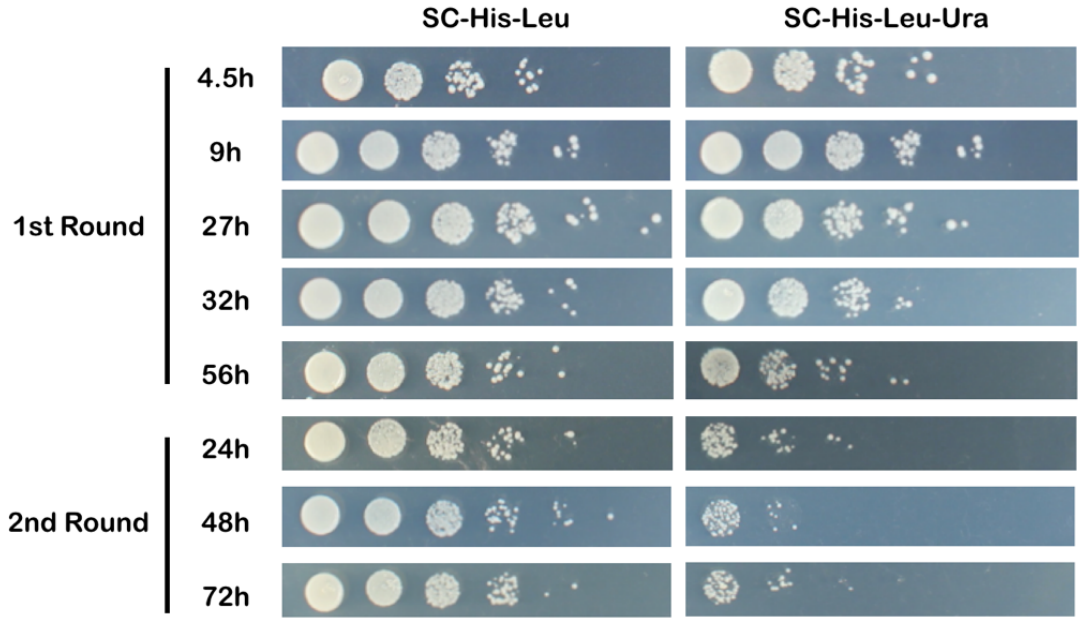


Figure 3.13| Recombination efficiency comparison with varying induction time. In the first round of induction, samples with P_{SCW11}DreEBD were spotted at 4.5h, 9h, 27h, 32h and 56h. A second round of induction was performed in parallel from 32h and continued for another 24h, 48h and 72h. The induction effect started to show after 27h induction; colony number continued to drop with longer induction as shown in 32h induction spotting; colony number of continued 1st round 56h induction is not as low as 2nd round 24h (still 56h in total) induction, indicating better induction with fresh estradiol after 32h; colony number continued to drop till 48h induction in the second round; no great colony number difference is observed between 48h and 72h induction in the second round.

In additional to timing optimization, a higher concentration of estradiol and a modified DreEBD fusion protein were tested for estradiol induction. The reasons why DreEBD is less efficient than CreEBD on induction could be that the fusion of Dre and the EBD generated spatial hindrance. On one hand, Dre might affect estradiol binding with the EBD or make it harder to trigger a conformational change to detach from Hsp90; on the other hand, four recombinase units have to work together on DNA as dimers, and when fused with EBD, Dre might be hindered in forming the quadruple-unit and become less efficient in driving DNA recombination. Therefore, a higher hormone concentration could help to increase the chance of effective hormone binding and release Dre from cytoplasm to nucleus. Increasing the length of linker of the fusion protein or changing linkers to turn EBD to a different angle could help to reduce spatial hindrance of EBD on Dre function. With the help of Dr. Jon Marles-Wright, the structure of DreEBD was simulated with individual model of Cre and EBD. Nonetheless, the method is limited since the structures of Dre and DreEBD were not available and the simulation can only provide clues. One possibility was that turning the angle of EBD might help to reduce mutual hindrance and the amino acid glycine (G) and aspartic acid (D) were inserted at the joining location of Dre and EBD to generate a new fusion version P_{SCW11}-Dre (+GD) EBD with a 5 aa linker (GDLEP). For the hormone concentration test and new fusion protein test, two rounds of liquid induction were performed and the second round induction was done after 48h induction of the first round. The result of spotting assay at 24h in first round induction and 48h in second round induction are shown in figure 3.14. In the 24h first round induction, strains with P_{SCW11}DreEBD and P_{SCW11}Dre (+GD) EBD have 10-fold reduction on triple dropout media compared to the double dropout with both 1μM and 10μM β-estradiol induction; in the 48h second round induction, strains induced by 1μM β-estradiol have roughly 10^3-fold colony number reduction and the P_{SCW11}Dre (+GD)EBD strain has slightly greater reduction than the P_{SCW11}DreEBD strain; strains induced by 10μM β-estradiol have roughly 10^4-fold colony number reduction and P_{SCW11}Dre(+GD)EBD also have slightly greater reduction than P_{SCW11}DreEBD (figure 3.14). The result illustrates that increased concentration of β-estradiol shows a slight

advantage with longer time induction over the lower concentration and modified DreEBD fusion protein also increased the recombination efficiency with longer time induction. Though the two methods did not generate a huge improvement of the recombination efficiency, it indicates that increasing inducer concentration and engineering of the linker region can potentially improve the recombination efficiency. But since the induction is very stringent and the recombination efficiency is sufficient for now, plus considering that the final purpose is to achieve moderate SCRaMbLE for the tRNA Neochromosome, it is not necessary to further increase induction efficiency until more information on the recombination of tRNA Neochromosome is obtained.

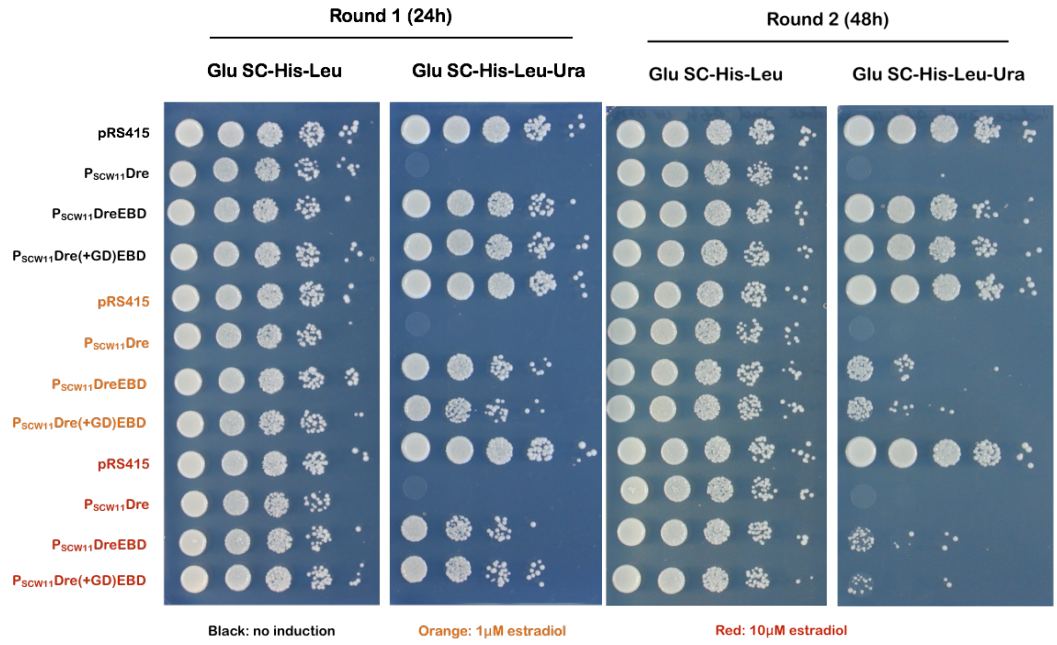


Figure 3.14| Induction result with different β-estradiol concentrations and a modified version of DreEBD. Both 1μM and 10μM β-estradiol were used for induction. A P_{SCW11}Dre (+GD) EBD modified fusion version was added for testing. The second round of induction was performed after 48h induction in the first round. In the 1st round 24h induction spotting assay, strain induced with 1μM and 10μM β-estradiol have similar 10-fold colony number reduction for both modified and non-modified DreEBD; in the 2nd round 48h induction, strains induced by 10μM estradiol had around 10^4-fold reduction, higher than those by 1μM estradiol which had 10^3-fold colony number reduction.

3.2.4.3 Function test of DrePBD and CrePBD

The function for CreEBD and DreEBD was demonstrated, and the next step is to develop another hormone-inducible recombinase for orthogonal SCRaMbLE. Cre and Dre have previously been fused with progesterone-binding domain (PBD) and functioned in mammalian cells (Anastassiadis et al., 2009; Wunderlich et al., 2001) and it was therefore decided to test whether this system is still functional in yeast. The induction molecule for the PBD is a progesterone-like chemical, RU486. In this experiment, both Dre and Cre were fused with PBD with the same 3 aa linker (LEP) in and induced by RU486. The replica plating assay shows that PBD can restrain Dre from working on rox sites in the absence of RU486 but can barely be released by RU486 (Figure 3.15a). The spotting assay with 48h plate induction is consistent with the replica plating result; no difference can be detected with or without RU486 induction (Figure 3.15b). Therefore, the first attempt to obtain the functional DrePBD failed. Fortunately, from the function test of CrePBD we can see that Cre can be successfully released and drive recombination to knock out *URA3* (Figure 3.16). The induction was performed for 48h in one round of induction and the spotting assay shows significant colony reduction on the triple dropout medium.

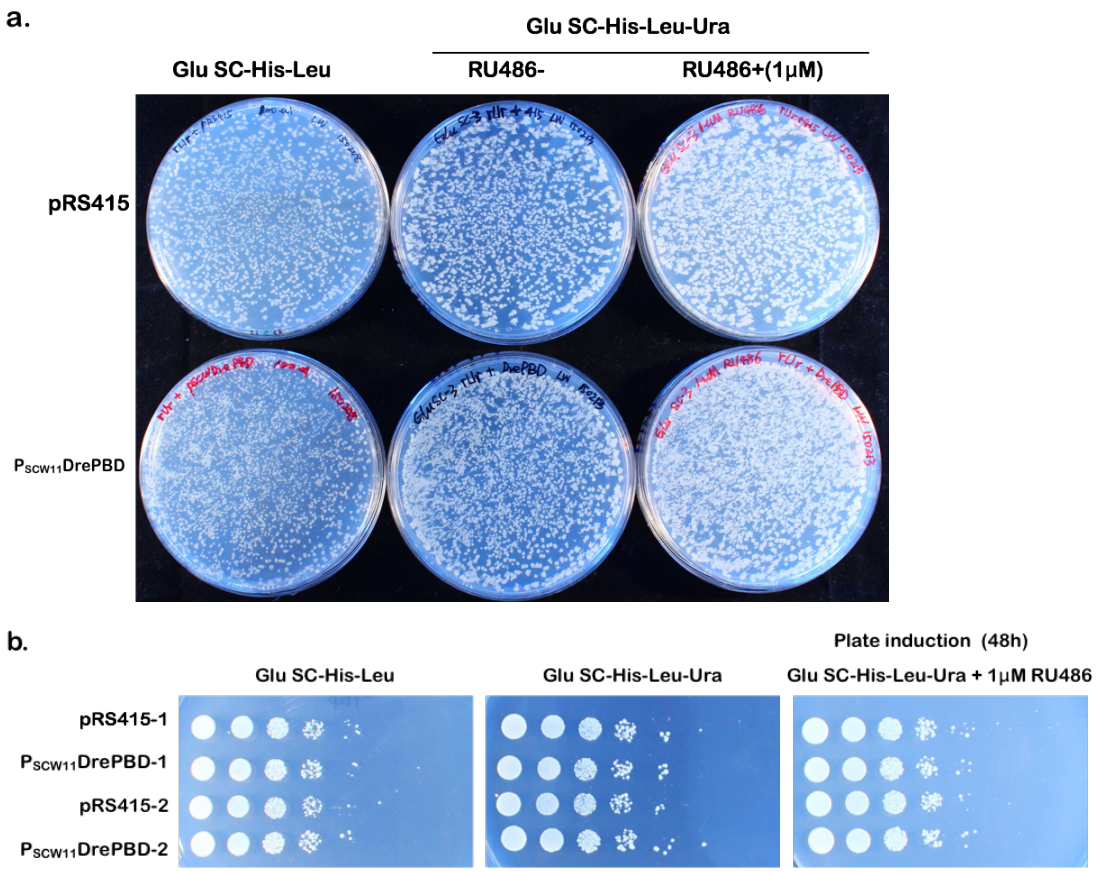


Figure 3.15| Function test of DrePBD. a. Replica plating result. Not much colony difference was detected on SC-His-Leu and SC-His-Leu-Ura plates with or without RU486. **b.** Spotting assay by plate induction. Not much colony difference was detected on SC-His-Leu and SC-His-Leu-Ura plates with or without RU486. DrePBD is not functional with RU486 induction.

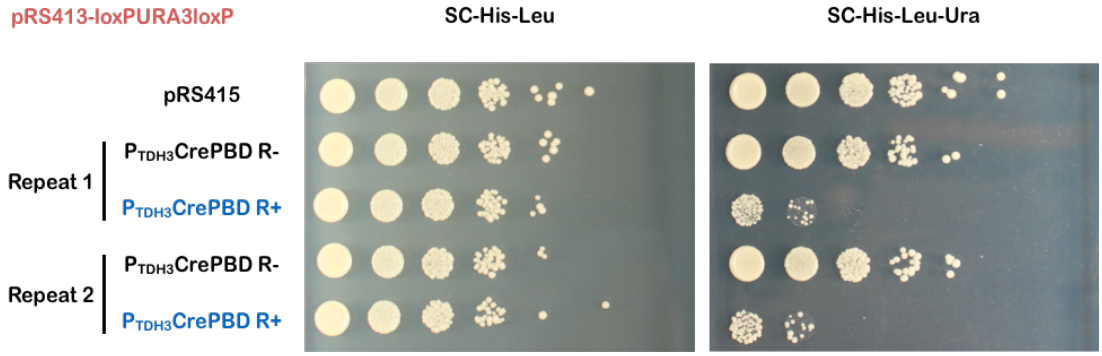


Figure 3.16| Function test of CrePBD. Spotting result of 48h liquid induction with RU486 on CrePBD shows that: on SC-His-Leu-Ura plate, there was an around $5X10^2$-fold colony number reduction with P_{TDH3}CrePBD, while there was not much difference with control vector pRS415 or without RU486 induction. The result shows that CrePBD is functional with RU486 induction. R-: no RU486 induction; R+: 1μM RU486 induction.

3.2.5 Orthogonality test for hormone regulation of Dre and Cre

The basic function of Dre and its orthogonality with Cre have been demonstrated. Two fusion proteins, DreEBD and CrePBD, have been designed and proved to be functional with β-estradiol and RU486 induction respectively. A final step is to see whether the inductions of the two molecules are orthogonal before moving on to SCRaMbLE the synthetic chromosomes. Strains with the two reporter devices pRS413-loxPURA3loxP and pRS413-roxURA3rox were further transformed with P_{TDH3}DreEBD, P_{TDH3}CrePBD and pRS415 control plasmid. After 48h induction of either 1μM β-estradiol or 1μM RU486, all strains were spotted on SC-His-Leu and SC-His-Leu-Ura glucose plates.

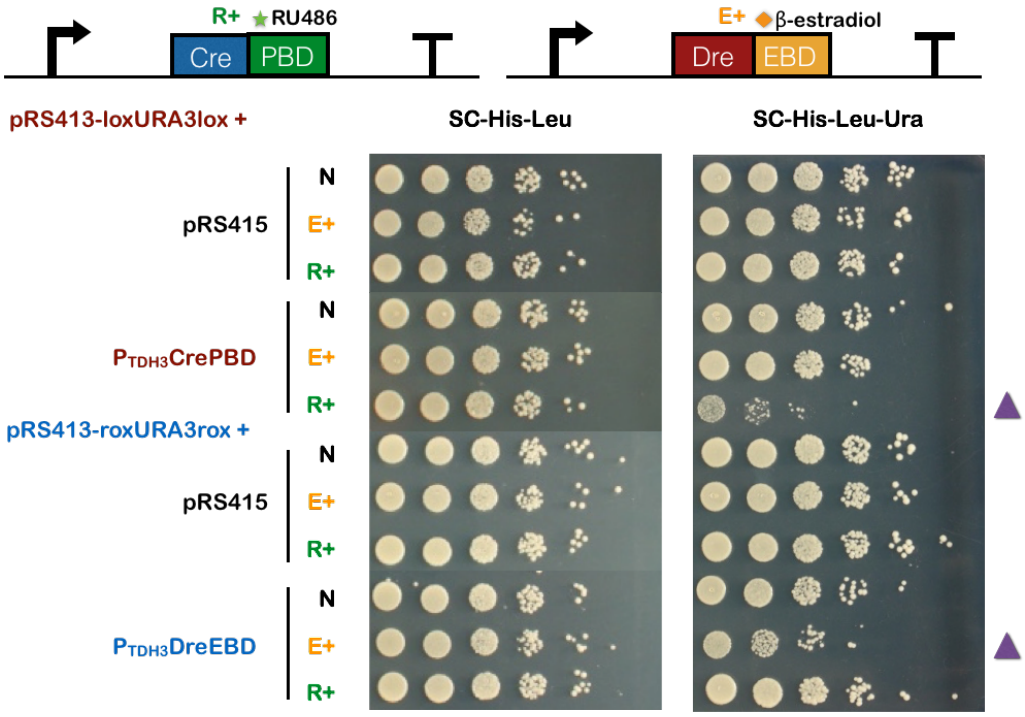


Figure 3.17| Orthogonal induction test of CrePBD and DreEBD. All strains were induced by either 1μM β-estradiol or 1μM RU486 in liquid medium for 43h. The spotting assay shows that only the strain with loxPURA3loxP and P_{TDH3}CrePBD induced by RU486, the strain with roxURA3rox and P_{TDH3}DreEBD induced by β-estradiol showed colony reduction on the SC-His-Leu-Ura plate. The result proves there is no cross talk between recombinase function and hormone induction for CrePBD and DreEBD. N: no induction; E+: 1μM β-estradiol; R+: 1μM RU486.

The spotting result shows that only the strain with loxPURA3loxP and P_{TDH3}CrePBD under the induction of RU486 shows colony number reduction on the triple dropout plate and only the strain with roxURA3rox and P_{TDH3}DreEBD under the induction of β-estradiol shows colony number reduction on the triple dropout plate (Figure 3.17). This demonstrates that CrePBD and DreEBD can function and be induced orthogonally by two different hormone molecules to drive DNA recombination and the toolkit can therefore be applied for orthogonal SCRaMbLE of the yeast synthetic chromosomes in the Sc 2.0 project.

3.2.6 Function test of three other back-up recombinases

3.2.6.1 Basic function test of Vika, SCre and VCre

The above sections focused on the functional characterization of Dre. Here the function of other backup recombinases is introduced to explain why Dre is a better candidate for the orthogonal SCRaMbLE toolkit. The backup recombinases include Vika, SCre and VCre. Since the activity of the P_{GAL1} promoter is leaky, only recombinase expression devices containing the P_{SCW11} promoter were used for the function test. Using the same function testing assay, two out of the three candidates were shown to be functional in yeast. The spotting assay shows that the strain with the pRS413-voxUravox device and P_{SCW11}-Vika, and the strain with the pRS413-SloxPUraSloxP device and P_{SCW11}-SCre grew well on the SC-His-Leu plate but failed to grow on the SC-His-Leu-Ura plate, indicating recombination due to Vika and SCre, while the VCre strain grew well on both SC-His-Leu and SC-His-Leu-Ura plate, indicating that VCre is not functioning properly in yeast (Figure 3.18). However, this is not conclusive since VCre might be toxic to the yeast genome. When looking at the colonies transformed with P_{SCW11}VCre plasmid, it was observed that there were both large and small colonies while at the beginning only large colonies were used for the spotting assay. After plasmid recovery from the large colonies and sequence verification, it was found that the VCre expression cassette was absent, possibly as a result of homologous recombination. Therefore, the spotting assay of these colonies does not indicate whether VCre is functional or not. The size of colonies transformed with P_{SCW11}VCreEBD was normal and similar to that transformed with pRS415,

suggesting that EBD may have restricted VCre from going into the nucleus (Figure 3.19). This could mean that there are pseudo VCre recombination sites on the yeast genome and VCre thus generate toxicity to the cell. However, no further experiment was performed to test the hypothesis considering that the purpose was to select at least one candidate for orthogonal SCRaMbLE, and VCre is clearly not suitable for this application.

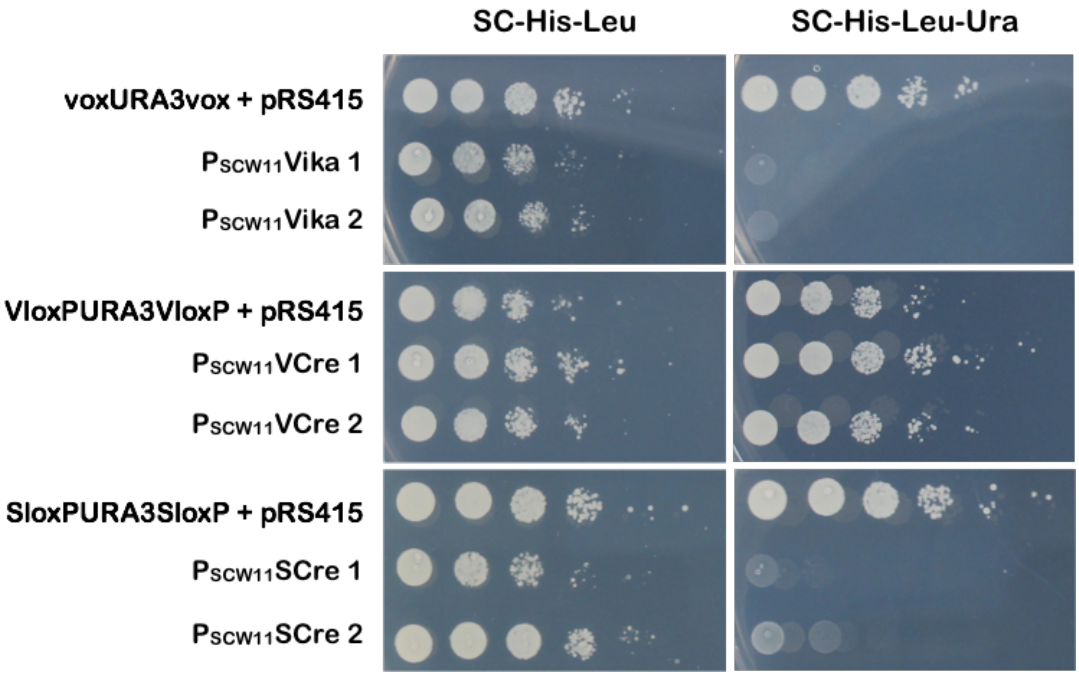


Figure 3.18| Basic functional test of Vika, VCre and SCre. Three reporter devices, VoxURA3Vox, VloxPURA3VloxP and SloxPURA3SloxP, were used for function testing. The spotting assay result shows that strains with P_{SCW11}Vika and P_{SCW11}SCre did not grow on the SC-His-Leu-Ura plate but that with P_{SCW11}VCre grew on SC-His-Leu-Ura plate as the control strain. Biological replicates (n=2) were performed for each recombinase test.

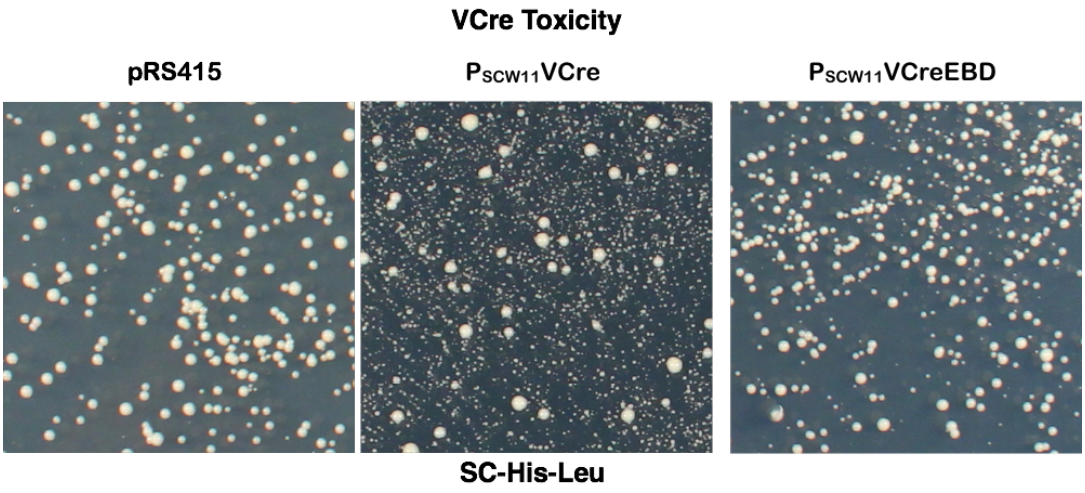


Figure 3.19|Colony size abnormality of strain with VCre expression device. Compared with the strain with control vector pRS415, the strain with P_{SCW11}VCre had mostly petite colonies and some colonies with normal size, indicating potential toxicity of VCre in yeast. The strain with P_{SCW11}VCreEBD had normal colony size, indicating that EBD somehow reduced the toxicity of VCre.

3.2.6.2 Induction test of recombinase EBD fusion

Since the basic function of Vika and SCre had been demonstrated, they were further fused with EBD for the hormone induction test. A spotting assay with plate induction was performed for P_{SCW11}VikaEBD and P_{SCW11}SCreEBD. The result shows that all cells grew well on an SC-His-Leu plate, while on a SC-His-Leu-Ura plate, with or without β-estradiol, cells failed to grow (Figure 3.20). This indicated that the EBD was not able to restrain the two recombinases from operating on DNA in the absence of β-estradiol. Taken the function test result of all recombinase and recombinase fusion proteins together, DreEBD works better than the rest of the candidate recombinases and can work together with CrePBD for the orthogonal SCRaMbLE application.

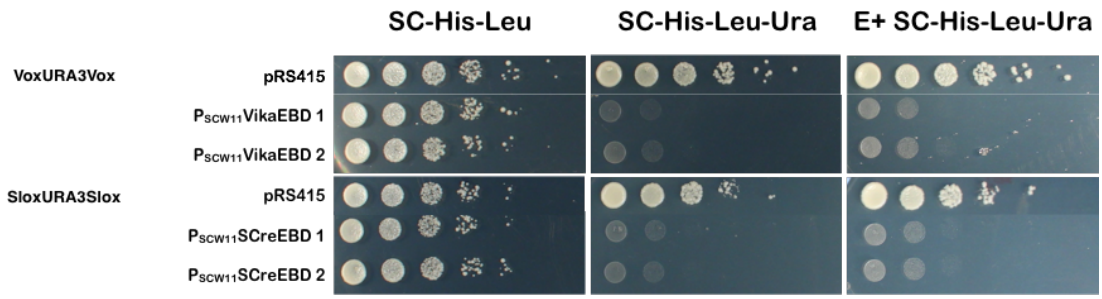
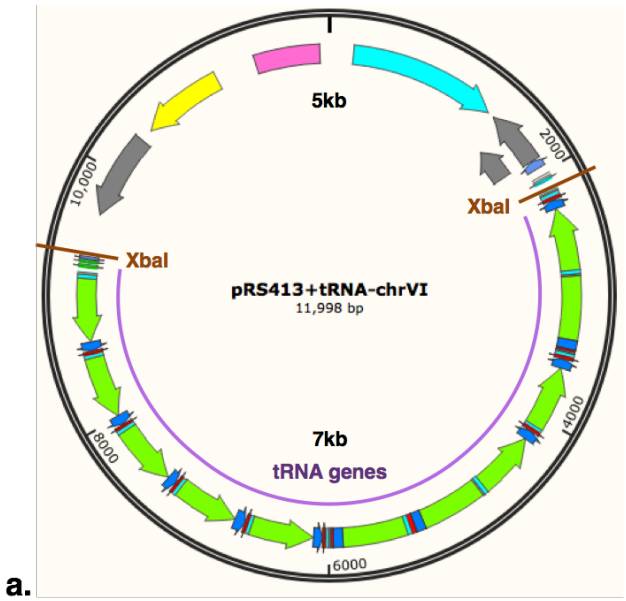


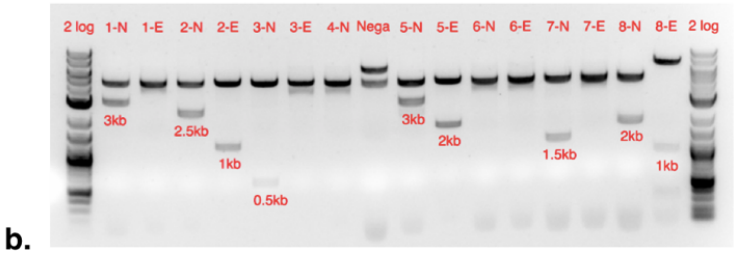
Figure 3.20|Plate induction spotting assay for Vika and SCre EBD fusion protein. The recombinase expression devices P_{SCW11}VikaEBD and P_{SCW11}SCreEBD were transformed to strains with respective reporter devices. All strains grew well on SC-His-Leu plates, but strains with VikaEBD or SCreEBD did not grow on SC-His-Leu-Ura plates, with or without β-estradiol induction. Biological replicates (n=2) were done for the assay. E+: 1μM β-estradiol.

3.2.7 SCRaMbLE of tRNA array by DreEBD

Since at the time when the recombinase toolkit was ready, the complete synthesis and characterization of the tRNA Neochromosome was not yet finished, the SCRaMbLE by DreEBD could not be performed on the whole tRNA Neochromosome. In the process of the construction of the Neochromosome, some partially synthesized tRNA arrays for different chromosomes were designed and constructed by another PhD student, Roy Walker. I was provided with the pRS413-tRNA-chrVI array which contains 10 tRNA genes from chromosome VI and one from chromosome V (Figure 3.21a). The pRS413-chrVI-tRNA array was first transformed into BY4741, followed by a second transformation with pRS415-P_{SCW11}DreEBD. The double transformed strain was cultured in SC-His-Leu liquid medium overnight and re-inoculated to OD_{600} of 0.1 into two tubes, one with a 7h induction of 1μM β-estradiol and the other un-induced. After the induction, cells were plated on SC-His plates and single colonies were picked for plasmid recovery. Since there are only two XbaI sites outside the tRNA gene clusters, the recovered plasmids were digested with XbaI to separate the tRNA genes (7028bp) from the backbone (4970bp) and the digestion map can indicate the changes of tRNA gene number. The digestion map of the negative control strain with pRS415 but not the Dre expression device shows 7Kb and 5Kb bands, indicating an intact tRNA cluster; while for those with P_{SCW11}DreEBD, there was a 5Kb backbone

band but the size of the other band varies from 0 to 3Kb, indicating partial or complete loss of tRNA genes (Figure 3.21b). Afterwards, all of the digestion tested samples were sent for sequencing to get the full genotype information of the remaining tRNA genes. The sequencing result is listed in figure 3.21c: green indicates that the tRNA is present while red means it is gone (Figure 3.21c). The sequencing result is consistent with the digestion result and random loss of tRNA genes was observed. However, it should be noted that with or without estradiol induction, there is tRNA loss. Four out of seven induced isolates lost all tRNA genes and two out of seven non-induced isolates lost all tRNA genes. In the previous function test with *URA3* reporter flanked by two rox sites, the function of DreEBD is tightly controlled; but here for the tRNA array the leakiness of DreEBD expression showed up indicating that the tRNA array is highly sensitive to Dre expression. There are several reasons that could explain this phenomenon: one is that there are more rox sites in the tRNA array and the possibility of Dre-rox binding is higher; another possibility could be that tRNA can be transcribed heavily and the frequency of open DNA duplex is higher, which could result in easier recombination by Dre; in addition, the recombination test was done in wild-type strain BY4741 but not the synthetic strains which were knocked out for the tRNA genes, therefore there is no selection pressure on the tRNA genes and they could be easily lost. It is not clear at present how it will work on the fully synthesized tRNA Neochromosome since the number of rox sites in the Neochromosome will be more than 270 and the number of rox sites of this 11-tRNA array could be much below the saturation point of Dre; furthermore, the tRNA Neochromosome will be put into a synthetic strain with selection pressure for essential tRNA genes; additionally, the tRNA array here is on a circular plasmid whereas the final tRNA Neochromosome will be linearized; the DNA form could also make a difference to the recombination rate. More timing and inducer concentration test should be done on the fully synthesized tRNA Neochromosome for SCRaMbLE.





c.

	1-N	1-E	2-N	2-E	3-N	3-E	4-N	5-N	5-E	6-N	6-E	7-N	7-E
tRNA V-20													
tRNA VI-1													
tRNA VI-2													
tRNA VI-3													
tRNA VI-4													
tRNA VI-5													
tRNA VI-6													
tRNA VI-7													
tRNA VI-8													
tRNA VI-9													
tRNA VI-10													

Figure 3.21| SCRaMbLE test of a tRNA gene array. a. Plasmid map of ChrVI tRNA array. The tRNA array was composed of 10 tRNA genes from chromosome VI and one from chromosome V and assembled into a pRS413 based vector. There are 11 tRNA genes and 12 rox sites in this array with all tRNA genes flanked with two rox sites. **b.** XbaI digestion pattern of isolates with SCRaMbLEed circular tRNA array. The length of the backbone is 5Kb and the intact tRNA cluster is 7Kb. The digestion patterns of the samples include one 5Kb backbone band and a 2nd band of different sizes, indicating various loss of tRNA genes, except for 8-E which is the digestion of P_{SCW11}DreEBD plasmid. Nega: negative control strain with tRNA array and pRS415 vector. N: non-induced; E: estradiol induced. **c.** Sequencing analysis of SCRaMbLEed circular tRNA array. Green: tRNA gene remains; Red: tRNA gene is lost. The loss pattern of the tRNA genes are all different and random. Among induced samples, 1,3,6,7 lost all tRNAs; among non-induced samples, 4 and 6 lost all tRNA genes.

3.2.8 SCRaMbLE of Synthetic chromosomes in haploid strains

In this section, the application of recombinase for SCRaMbLEing yeast synthetic chromosomes will be introduced. Before using the engineered CrePBD designed and constructed in this study, the published CreEBD device (Lindstrom & Gottschling, 2009) was first used for function testing on synthetic chromosomes. The first partially synthesized right arm of chromosome IX (SynIXR, 91Kb with 43 loxPsym sites) (Dymond et al., 2011), and the total synthesized chromosome II (SynII, 770Kb with 267 loxPsym sites) (Shen et al., 2017) were used for Cre SCRaMbLE. Since during the SCRaMbLE process, random loss of genes, including essential genes, can result in cell death and affect the survival rate of the whole population, serial dilution and the spotting assay can be used for CreEBD function and induction verification. Both plate induction and liquid induction were performed for SCRaMbLE SynIXR and SynII strains, and BY4741 was used as a negative control. The induction time for plate induction was 48h and for liquid induction was 24h. The spotting result shows that BY4741 strain with CreEBD grew well with or without estradiol induction, showing no effect of CreEBD expression on cell growth of the wild-type strain; SynIXR and SynII strains with CreEBD grew well without estradiol induction but have obviously lower survival rate with estradiol induction, with lower survival rate in the case of plate induction (Figure 3.22).

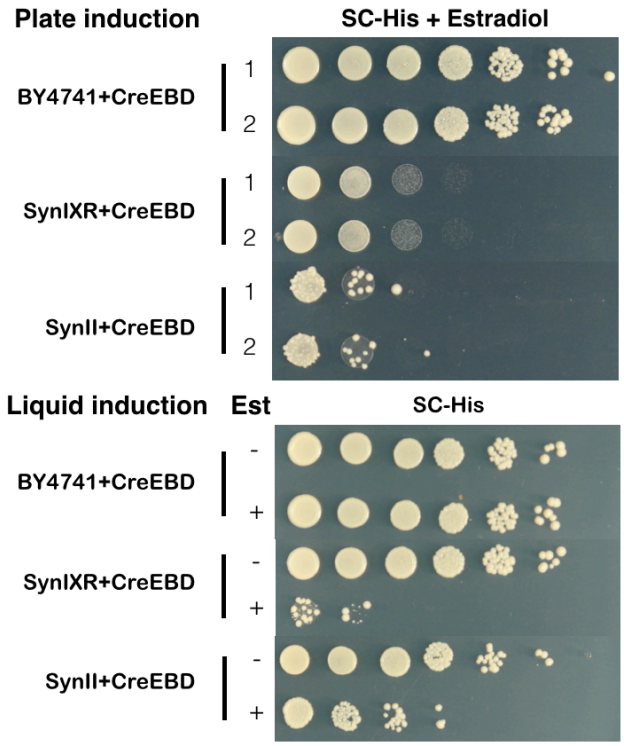


Figure 3.22| SCRaMbLE function test of CreEBD on the strains with synthetic chromosomes. BY4741, SynIXR and SynII were used to test SCRaMbLE function of CreEBD. In plate induction, BY4741 with CreEBD grew well while SynIXR and SynII with CreEBD all have much lower survival rate; in liquid induction, BY4741 with CreEBD grew well with or without estradiol induction; SynIXR and SynII grew well without induction but have much fewer survivors with estradiol induction.

3.2.9 SCRaMbLE of multiple synthetic chromosomes for pathway diversification

In this section, both the SCRaMbLE function test of CrePBD and primary experiments on SCRaMbLE of the synthetic yeast chassis for pathway diversification will be introduced. In addition to the CrePBD SCRaMbLE test, a violacein- or β-carotene-producing pathway was also introduced into the synthetic strain. Violacein is a violet pigment that has antitumor and antibiotic activities. There are five genes involved in the pathway, *VioA, B, D, C, E*. In the process of violacein production, some green intermediates, prodeoxyviolacein and proviolacein, are produced before the final transformation to the purple violacein (Lee et al., 2013). Therefore, the production stage of the reaction can be estimated from the color of a yeast producer. β-carotene is an orange pigment commercialized as food supplement. For the production in

yeast, three genes were introduced including *CrtI, CrtE, CrtYB* (Verwaal et al., 2007). Before the final transformation to β-carotene, two intermediates are generated, including a yellow pigment Neurosporene and a red pigment Lycopene. The chemical composition can also be determined from the colony color. Therefore, the violacein and β-carotene pathways were used as demonstration pathways for diversification with SCRaMbLEed synthetic strains. Mitchell's work was referred to for the expression of the two pathways in yeast (Table 3.4) (Mitchell et al., 2015). With the complete synthesis of more chromosomes, some synthetic chromosomes were combined together by collaborators from New York University, and Tsinghua University, including SynIII-VI-IXR and SynII-XII. We were also gifted with a SynV chromosome from Tianjin University. These new strains with different synthetic chromosomes were explored for pathway diversification (Table 3.5).

Table 3.4 Genetic information for β-carotene and violacein pathways

Promoter	**ORF**	**Terminator**
P_{TDH3}	CrtI	T_{ACS2}
P_{PGK1}	CrtE	T_{ACS1}
P_{ACT1}	CrtYB	T_{ENO2}
P_{TDH3}	Violacein A	T_{ACS2}
P_{PGK1}	Violacein B	T_{ACS1}
P_{ACT1}	Violacein C	T_{ENO2}
P_{RPS2}	Violacein D	T_{CIT1}
P_{ZEO1}	Violacein E	T_{FUM1}

The SynII-XII combined synthetic strain LWy078 was chosen for SCRaMbLE function test of CrePBD. The violacein expressing circuit was put on the pRS416 vector and transformed into the synthetic and the wild-type control strains. BY4741 strain with P_{TDH3}CrePBD-pRS415 and the SynII-XII strain with pRS415 were used as controls. From the 24h and 48h RU486 liquid induction spotting result, we can see that only SynII-XII with P_{TDH3}CrePBD has obviously low survival rate on the SC-Leu-Ura plate,

indicating the function and induction of CrePBD (Figure 3.23). We can also tell from the assay that the colony size of the wild-type strain with the violacein pathway is larger than that of the synthetic strain, indicating that the pathway could create more burden on the synthetic strain.

Table 3.5 Table of synthetic strains

Strain ID	Description	Genotype	Source
LWy095	SynII 1.3	*leu2Δ0 met15Δ0 LYS2 ura3Δ0 his3Δ1*	Cai Lab (UoE)
LWy096	SynIII-VI-IXR	*leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII SYN.SUP61::HO synVI SYN-WT.PRE4 IXL-synIXR*	Boeke Lab (NYU)
LWy078	SynII-XII	*MATa his3Δ1 leu2Δ0 ura3Δ0 MET15 + LYS2 synOJ::HIS3*	Dai Lab (THU)
LWy119	SynV	*MATa His3Δ1 leu2Δ0 met15Δ0 ura3Δ0 LYS2 synV*	Yuan Lab (TJU)

Besides chemical production diversification, SCRaMbLE can also be applied to screen for a chassis with lower metabolic burden of the heterologous pathway. During the RU486 induction, the cells were also plated at certain time points to screen for colonies that have color variations. The candidate colonies were further streaked out on SC-Ura plates for color comparison. After 3 days of growth, the colonies were compared between control strains and the SCRaMbLEed strains at different induction time points (Figure 3.24). The control strains include BY4741 with CrePBD, SynII-XII with pRS415 and SynII-XII with CrePBD but without RU486 induction. The color of BY4741 with violacein is darker than SynII-XII and there is not much difference between the SynII-XII control strains. For the SCRaMBLEed strains, some have lighter color and some have darker color than the SynII-XII control strains despite the induction time difference. The colony size was also more variable after SCRaMbLE.

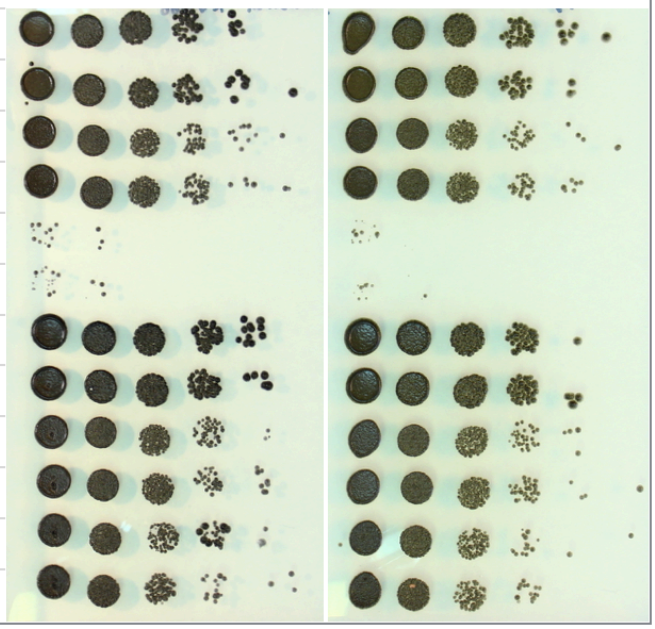
Strain	Plasmid	RU486	Replicate	24h	48h
BY4741	pTDHCrePBD	+	1		
BY4741	pTDHCrePBD	+	2		
SynII-XII	pRS415	+	1		
SynII-XII	pRS415	+	2		
SynII-XII	pTDHCrePBD	+	1		
SynII-XII	pTDHCrePBD	+	2		
BY4741	pTDHCrePBD	-	1		
BY4741	pTDHCrePBD	-	2		
SynII-XII	pRS415	-	1		
SynII-XII	pRS415	-	2		
SynII-XII	pTDHCrePBD	-	1		
SynII-XII	pTDHCrePBD	-	2		

Figure 3.23| SCRaMbLE function test of CrePBD for SCRaMbLE of a yeast strain with two synthetic chromosomes and the violacein pathway. BY4741 with P_{TDH3}CrePBD-pRS415, SynII-XII strain with pRS415 and SynII-XII with P_{TDH3}CrePBD-pRS415 were used for the induction and spotting assay. The former two were controls. Cells were spotted to SC-Ura-Leu plates after 24h and 48h RU486 induction respectively. Only SynII-XII with P_{TDH3}CrePBD under RU486 induction showed reduced survival after induction.

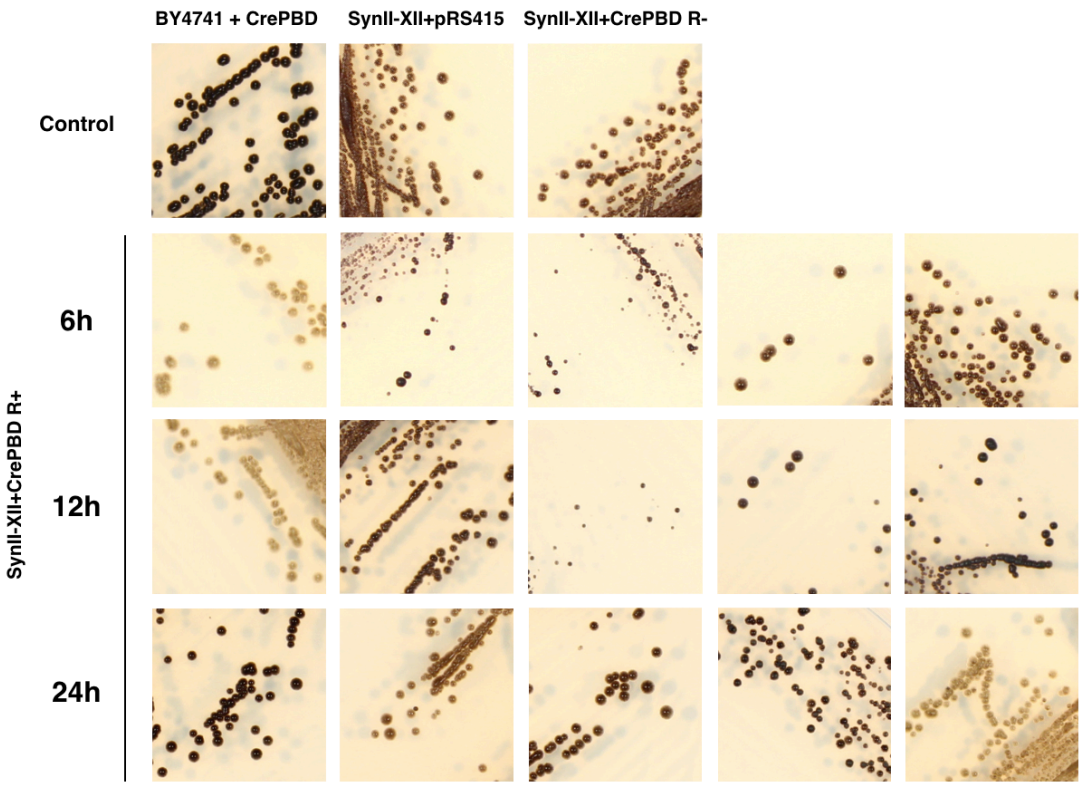


Figure 3.24| Diversified colony color and size of SynII-XII violacein SCRaMbLEed strain. SCRaMbLEed samples were plated on SC-Ura after 6h, 12h and 24h RU486 induction and re-streaked for single colony comparison. Different colony colors, both lighter and darker, can be observed compared with the control strains. The colony size was also diversified after SCRaMbLE. Control: BY4741 with CrePBD, SynII-XII with pRS415, SynII-XII with CrePBD but no induction.

Similarly, the β-carotene pathway was also put into SynII, SynV, SynII-XII and SynIII-VI-IXR synthetic strains for SCRaMbLE. Since the color difference between yellow and orange is not as distinctive, candidate colonies were selected and spotted on plates for color comparison after SCRaMbLE. The spotting shows that the extent of color variation is different among these SCRaMbLEed synthetic strains (Figure 3.25). Compared with the control strain, more obvious difference is observed in SCRaMbLEed SynV strains, indicating that β-carotene diversification is not sensitive to SCRaMbLE of all synthetic strains, but tends to be more sensitive to the SCRaMbLE of SynV in this case. It could be that there are more relevant genes which can affect β-carotene production on chromosome V but it could also be related to the efficiency and randomness of SCRaMbLE. Therefore, a screening strategy should be developed

to give a selection pressure to obtain the SCRaMbLEed strains more efficiently. A SCRaMbLE selection strategy will be introduced in the next chapter followed by deeper chemical analysis and genotype analysis on the SCRaMbLEed strains.

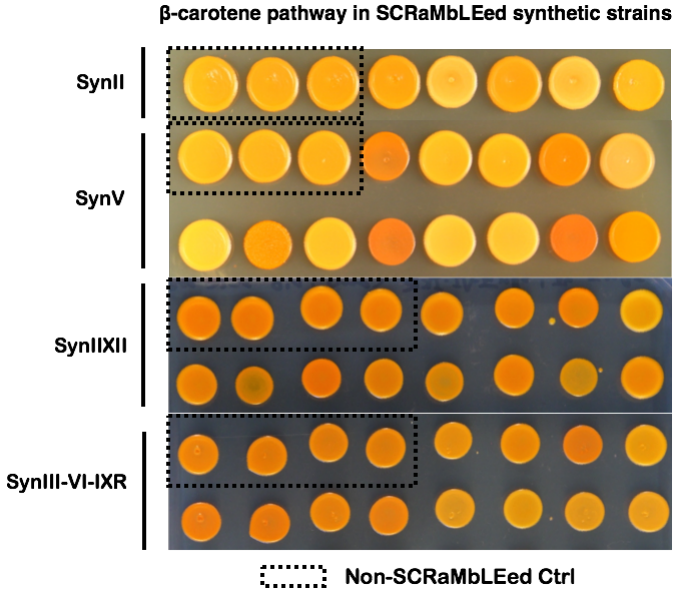


Figure 3.25| Color comparison between SCRaMbLEed synthetic strains with β-carotene synthesis pathway. Single colonies from SCRaMbLEed samples were inoculated to liquid media and standardized to OD_{600} of 1.0 before spotting. Spots in the dashed boundaries are non-SCRaMbLEed controls. The range of color difference is larger for the pathway in SynV after SCRaMbLE.

3.3 Chapter summary, discussion and future work

This chapter focused on the functional characterization of five tyrosine recombinases and their application for SCRaMbLE of synthetic chromosomes in yeast. Among the five candidate recombinases, Cre and Dre were proven to be the best two recombinases for SCRaMbLE and their recombination functions are orthogonal. Cre and Dre were further engineered by fusion with hormone binding domains from different nuclear receptors for precise function activation. The fusion protein DreEBD can be induced by β-estradiol and CrePBD can be induced by RU486, with no cross talk in activation. The final aim of the recombinase toolkit is to achieve orthogonal SCRaMbLE of the tRNA Neochromosome and the other synthesized chromosomes. Since the tRNA Neochromosome was not fully synthesized and has not been characterized, the function of DreEBD was only tested on a partially synthesized tRNA

array. Though more induction optimization should be done for the tRNA Neochromosome in future, DreEBD was proved to generate random tRNA loss in the tRNA array. For the SCRaMbLE of normal synthesized chromosomes, CrePBD was used to SCRaMbLE SynIXR, SynII, SynV, SynII-XII and SynIII-VI-XI. Pathway diversification was initially tested by SCRaMbLEing these synthetic strains with two demonstration pathways: violacein and β-carotene synthesis pathways. Now, the synthesis of the tRNA Neochromosome is almost done and waiting to be characterized; the synthesis of other synthetic chromosomes is almost half complete. The next step for the project is to achieve co-SCRaMbLE of the tRNA Neochromosome and the other combined synthetic chromosomes for different applications such as generating a minimal strain with reduced genome size, faster growing and robust industrial strains for fermentation, improving the yield of high value natural products and biofuels, etc. The characterization of the SCRaMbLEed strain with pathways of interest will be introduced in the next chapter.

Chapter 4 Chassis level pathway diversification in synthetic yeast

4.1 Overview and circuit design

After the orthogonal recombinase toolkits were set up, further work has been focused on the SCRaMbLE of synthetic chromosomes by Cre/loxP. Since *S. cerevisiae* is one of the most common species for fermentation and industrial application, the main interest is in using the synthetic yeast and the SCRaMbLE system to improve heterologous natural product production by chassis diversification and optimization. The main strategy for optimizing heterologous pathways includes fine-tuning the expression level of each gene in the pathway, while the genomic context of the heterogeneous chassis is also of vital importance. Local host pathways can compete with heterologous pathways for precursors and energy molecules. Lowering gene expression in competing pathways or deleting genes interacting with the target pathway can re-direct precursor fluxes or remove toxicity of side products to achieve higher target production. However, deleting local genes could also affect the fitness of the host cell. Simultaneous whole genome optimization has the potential to overcome such impact. The genome editing tool CRISPR/Cas9 system can achieve editing of single or multiple genes, but it is not yet able to cover the possibility of gene rearrangements systematically on a whole genome level and may also suffer from off-target effects.

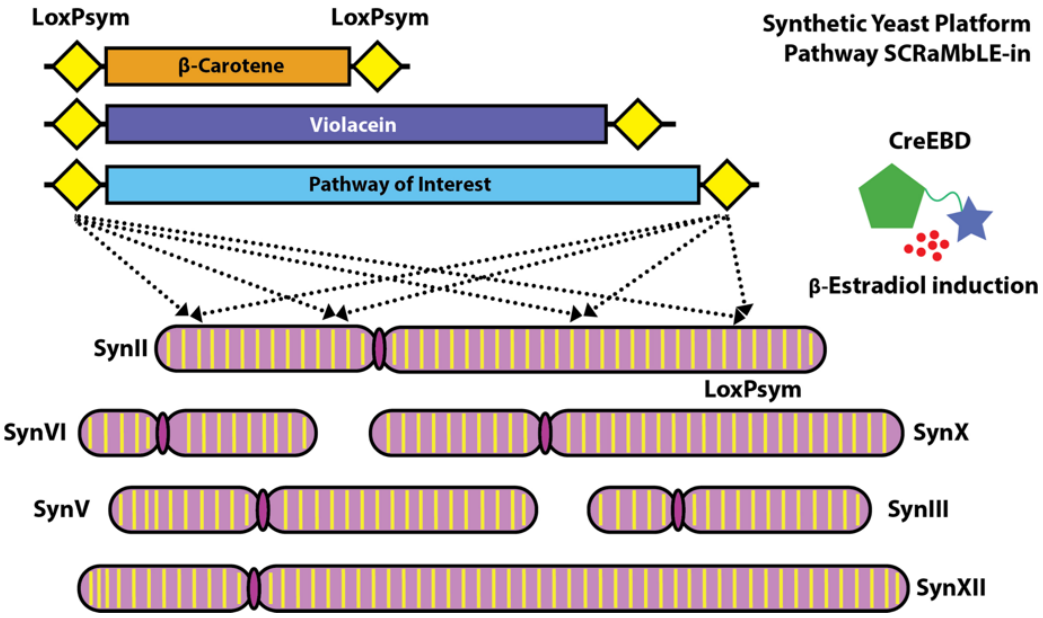


Figure 4.1| Overview of pathway integration and chassis level diversification in the synthetic yeast. In the Sc 2.0 synthetic yeast project, the design of inserting loxPsym sites in the synthetic chromosomes enables random genome rearrangement by the activation of Cre recombinase (SCRaMbLE). The synthetic yeast is used for chassis optimization of expression of heterologous pathways. A universal SCRaMbLE-in device was designed to integrate any pathway of interest into the synthetic chromosomes. The β-carotene and violacein synthesis pathways were used for demonstrating pathway integration by the SCRaMbLE-in device.

In the synthetic yeast genome Sc 2.0 project, the SCRaMbLE system, Synthetic Chromosome Recombination and Modification by LoxP-Mediated Evolution, was designed to introduce genome-wide LoxPsym sites and can generate theoretically nearly infinite gene order permutations (Dymond et al., 2011). This design was seen as the most intriguing and potentially valuable design change added to the synthetic genome (Jovicevic et al., 2014; Shen et al., 2016). With the continuing successful synthesis and construction of these synthetic chromosomes, their stable existence and minimal effect on phenotypes (Annaluru et al., 2014; Mercy et al., 2017; Richardson et al., 2017; Shen et al., 2017), the synthetic yeast can be a unique novel platform for metabolic engineering. In addition to the genomic composition, the location of integration could affect the transcription efficiency differently depending on genome structure and the degree of heterochromatinization (Ay et al., 2014;

Englaender et al., 2017; Kitada et al., 2012; Mano, Kobayashi, Nakayama, Uchida, & Oki, 2013). Also, the scattered cis-acting elements and other not fully characterized untranslated regions can also make a difference (Yoshikawa, Furusawa, Hirasawa, & Shimizu, 2008). Moreover, compared with being carried on a plasmid, the pathway integrated into the genome avoids the burden of an additional selection marker and favors stable replication and expression in the host. Though the effect of some locations in the genome has been characterized (Englaender et al., 2017; Flagfeldt, Siewers, Huang, & Nielsen, 2009), this characterization has not yet systematically covered all the locations in the genome. We therefore designed a universal "SCRaMbLE-in" device that can integrate target pathways randomly into any location between open reading frames in the genome and at the same time drive genome rearrangements (Figure 4.1). Through a corresponding screening method, integration and genome rearrangements can be analyzed for strains with special phenotypes. Synthetic chromosome II, which contains 410 open reading frames and 267 loxPsym sites, was chosen as the demonstration chromosome for pathway integration (Shen et al., 2017). The loxPsym site (ATAACTTCGTATAATGTACATTATACGAAGTTAT) is a symmetrical spacer mutant of loxP (Hoess et al., 1986) and when the SCRaMbLE process is activated, the pathway can be integrated randomly into the synthetic chromosome with simultaneous random gene deletion, inversion and duplication of genes on the chromosome, which is called a "SCRaMbLE-in" process.

4.1.1 Universal SCRaMbLE-in device

A universal SCRaMbLE-in device that can accommodate the element of interest was designed to achieve SCRaMbLE-in to any synthetic chromosome when the whole genome synthesis is completed. The SCRaMbLE-in device was constructed from a yeast centromeric plasmid to control the copy number, and includes a YeastFab compatible BsaI sites flanking an RFP cassette for pathway loading, a *LEU2* auxotrophic marker for integration selection between two loxPsym, sites and a *URA3* counter selection marker (Figure 4.2). When recombinase is expressed in yeast, the pathway is first excised from the device and then integrated into the synthetic

genome. A *URA3* counter-selection strategy was used for screening genome integrated strains. With the expression of *URA3*, the substrate 5-FOA can be converted to the toxic form (i.e., 5-fluorouracil). If recombination does not occur, plasmid-bearing cells will die because of the presence of the *URA3* gene; if recombination happens on the plasmid, the pathway with *LEU2* will be separated from the *URA3* gene, and only strains with the genome-integrated pathway with the loss of the *URA3* gene can survive on 5-FOA SC-Leu plates. Different genes, pathways or any element of interest can be assembled into the SCRaMbLE-in device for genome integration. To study the impact of integration and chassis diversification, three GFP expression cassettes, a β-carotene and a violacein synthesis pathway were used for testing.

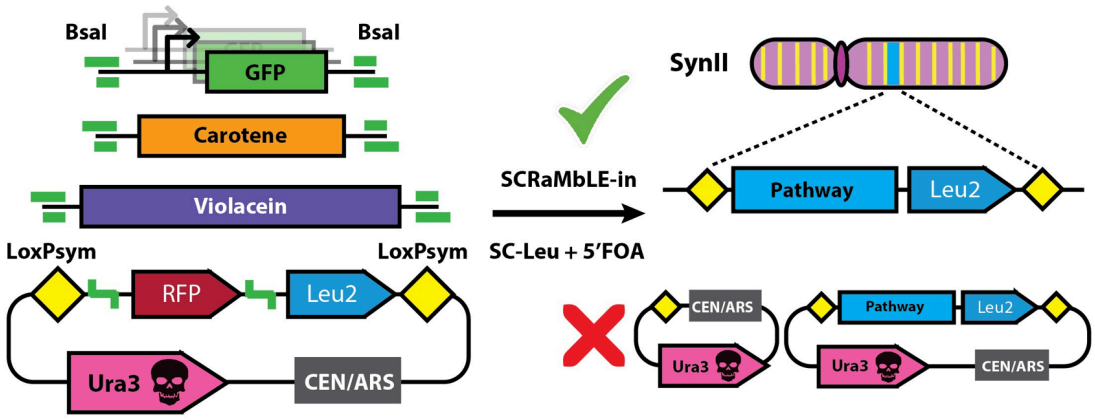


Figure 4.2|Design and selection strategy of the SCRaMbLE-in device. The SCRaMbLE-in device is based on a yeast centromeric plasmid with *URA3* marker. A BsaI site-flanked RFP cassette and a *LEU2* expression unit were put between two loxPsym sites. *LEU2* is used as positive selection marker for integration and *URA3* is used as counter selection marker for non-integrated strains. After SCRaMbLE-in induction, a SC-Leu with 5-FOA plate was used for pathway integration selection.

4.1.2 SCRaMbLE-in workflow

The workflow of SCRaMbLE-in includes induction, screening and characterization of genotype and phenotype of SCRaMbLE-in variants (Figure 4.3). The SCRaMbLE-in device is first transformed into the synthetic strain by SC-Leu selection and further transformed with a CreEBD expression plasmid by SC-His-Leu selection. After 12h, 24h or 48h of β-estradiol induction, the culture is plated to SC-Leu 5-FOA for selecting

genome integrated strains. A spotting assay is also performed on SC-Leu and SC-Leu 5-FOA to evaluate the efficiency of integration. Survivors on the counter-selection plate are potential SCRaMbLE-in candidates, but they need further verification by colony PCR of the *URA3* region, to ensure the absence of *URA3* rather than its inactivation by point mutation. The verified integrants are characterized either by flow cytometry for the GFP cassette integration or by chemical analysis for pathway integration. For those with interesting phenotypes, the integration location and whole genome variation are confirmed with next generation sequencing. Finally, the contributions of genotype variations to phenotype variations can be further analyzed and investigated. Genome integration is one aspect of SCRaMbLE-in; continuous SCRaMbLE is still necessary to generate more genome rearrangements for chassis optimization. Therefore, after the confirmation of genome integration, those with a higher level of GFP expression or higher chemical production can be continuously SCRaMbLEed to screen for strains with further optimized chassis.

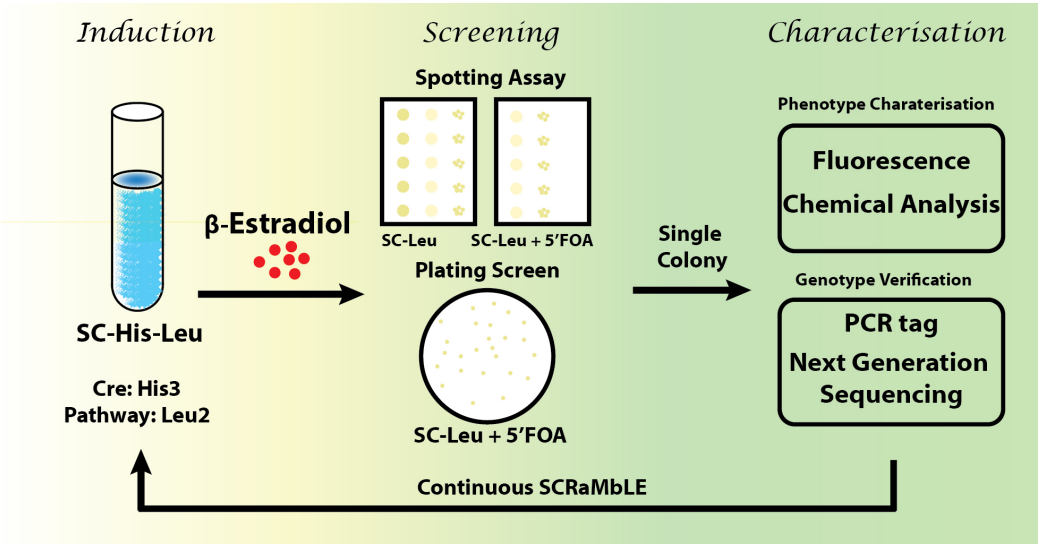


Figure 4.3|Illustration of SCRaMbLE-in workflow. The procedures for SCRaMbLE-in include induction, screening, characterization and continuous SCRaMbLE. Liquid medium SC-His-Leu was used for selection of strains with the pathway SCRaMbLE-in device and the Cre expression cassette. β-estradiol was added for induction. Spotting assay and plating screening were done on the 5-FOA counter-selection plates. Single colonies with special phenotypes were picked for further analytical characterization and verification of genome integration. Successful integrants were carried on for cycled SCRaMbLE and verification to be further diversified and screened for expected phenotypes.

4.1.3 DNA circuits and strains list

The DNA constructs used in this study are summarized in Table 4.1 and the SCRaMbLE-related yeast strains are listed in Table 4.2.

Table 4.1 DNA circuit list for SCRaMbLE-in

Strain ID	Plasmid ID	Description
LWe108	pWL032	Universal SCRaMbLE-in acceptor vector
LWe168	pWL045	β-carotene SCRaMbLE-in device
LWe177	pWL046	Violacein SCRaMbLE-in device
LWe983	pWL334	P_{TDH3}-GFP SCRaMbLE-in device
LWe984	pWL335	P_{ACT1}-GFP SCRaMbLE-in device
LWe985	pWL336	$P_{YGR270W}$-GFP SCRaMbLE-in device
LWe991	pWL341	pRS413-YBR044C
LWe992	pWL342	pRS413-YBR296C
LWe993	pWL343	pRS423-YBR044C
LWe994	pWL344	pRS423-YBR296C

Table 4.2 SCRaMbLEed strains

Strain ID	Strain	Description
LWy134	*SynII*	HO locus integrated β-carotene in *SynII* strain
LWy137	*SynII*	HO locus integrated violacein in *SynII* strain
LWy212	By4741	HO locus integrated β-carotene in BY4741
LWy213	By4741	HO locus integrated violacein in BY4741
LWy152	*SynII*	Violacein SCRaMbLE-in isolate 1
LWy238	*SynII*	Violacein SCRaMbLE-in isolate 2
LWy239	*SynII*	Violacein SCRaMbLE-in isolate 3
LWy256	*SynII*	LWy152+ generated from continued SCRaMbLE
LWy257	*SynII*	LWy152- generated from continued SCRaMbLE
LWy258	*SynII*	LWy238+ generated from continued SCRaMbLE
LWy259	*SynII*	LWy238- generated from continued SCRaMbLE

Strain ID	Strain	Description
LWy260	*SynII*	LWy239+ generated from continued SCRaMbLE
LWy261	*SynII*	LWy239- generated from continued SCRaMbLE
LWy215	*SynII*	β-carotene SCRaMbLE-in isolate 1
LWy252	*SynII*	β-carotene SCRaMbLE-in isolate 2
LWy253	*SynII*	β-carotene SCRaMbLE-in isolate 3
LWy220	*SynII*	SynII with SCR-in P_{TDH3}-GFP plasmid
LWy221	*SynII*	SynII with SCR-in P_{ACT1}-GFP plasmid
LWy222	*SynII*	SynII with SCR-in $P_{YGR270W}$-GFP plasmid
LWy232	*SynII*	SynII HO integrated P_{TDH3}-GFP
LWy233	*SynII*	SynII HO integrated P_{ACT1}-GFP
LWy234	*SynII*	SynII HO integrated $P_{YGR270W}$-GFP
LWy245	*SynII*	Chr15 *HIS3* locus integrated P_{TDH3}-GFP isolate after SCRaMbLE-in

4.2 Results

4.2.1 Preparation of automated screening method for SCRaMbLE

4.2.1.1 Fluorescence distribution of SCRaMbLEed strains with GFP expression

Before performing chassis optimization for specific pathways, the effect of SCRaMbLE on the expression of a fluorescent protein in the yeast with synthetic chromosome II (*SynII* strain) was evaluated. Three GFP cassettes with different promoters (500bp 5'UTR of the gene), P_{TDH3} (strong), P_{ACT1} (medium) and $P_{YGR270W}$ (weak) were used for the comparison. The three cassettes were assembled into a centromeric plasmid by golden gate assembly for expression in the synthetic strain. The fluorescence distribution result shows that the expression patterns of SCRaMbLEed strains are similar among the cell mixtures with the three different GFP expression cassettes: after SCRaMbLE, cell numbers with lower fluorescence intensity increase while those with higher fluorescence decrease (Figure 4.4). This phenomenon is expected since fragment deletion is the most common type of genome rearrangement and it makes sense that the fitness and gene expression status be diminished for most of the

population. However, previous data show that the SCRaMbLE system can recover ill-growing *SynII* on YPG media with glycerol as carbon source at 37°C (Shen et al., 2017). This is a stringent selection condition and only a very small population can be selected from the SCRaMbLEed population. Similarly, for the process of selecting SCRaMbLEed strains with higher heterologous gene expression, proper selection is also needed and for fluorescent proteins the obvious method is to apply fluorescence-activated cell sorting (FACS) for screening out the higher expressing strains.

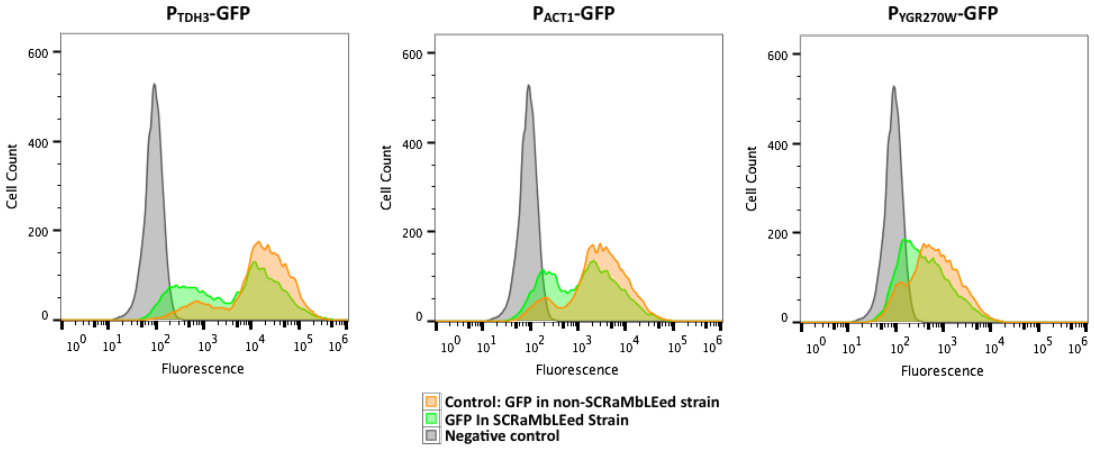


Figure 4.4| Fluorescence distribution of SCRaMbLEed GFP expressing strain. Three different GFP expression cassettes, P_{TDH3}-GFP, P_{ACT1}-GFP and $P_{YGR270W}$-GFP, were used to evaluate the effect of SCRaMbLE on heterologous gene expression. The result shows that after SCRaMbLE the fluorescence peak moves to left, indicating a higher percentage of cells with lower fluorescence intensity. The patterns are similar among the three different GFP expression cassettes. Negative control: *SynII* strain.

4.2.1.2 Selection of strains with higher GFP expression by FACS

To enable high-throughput screening of high GFP-expressing strains from a large population, the FACS method was tested on LWy245, a *SynII* strain with the P_{TDH3}-GFP cassette integrated to the *HIS3* locus at chromosome XV. Using the FACSAria cell sorter, the SCRaMbLEed population was gated by visible cells, intact cells, single cells, GFP expressing cells, High GFP (top 12% fluorescence intensity) and Low GFP (bottom 12% fluorescence intensity) cells (Figure 4.5a). The High GFP and Low GFP population were sorted and recovered on SC-Leu plates for fluorescence evaluation. The plating result shows that the sorted cells can be successfully recovered and there is indeed

fluorescence intensity difference after the sorting (Figure 4.5b). The fluorescence was roughly estimated by BIORAD imaging.

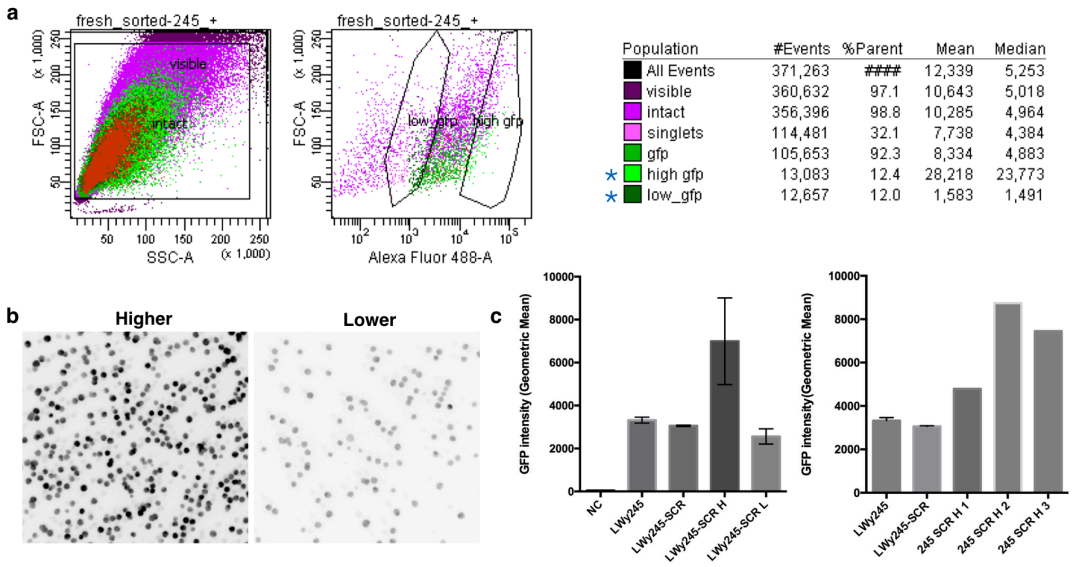


Figure 4.5|Cell sorting of GFP-expressing strain in the SCRaMbLEed yeast. a. Dot plot and gating of GFP expressing SCRaMbLEed yeast by flow cytometry. According to the FSC vs SSC plot, the cells were gated to visible particle, intact cells, single cells; according to the FSC vs Alexa Fluor 488 channel plot, the cells were further gated to GFP expressing cells, High GFP cells and Low GFP cells. The High GFP and Low GFP cells were selected from the top and bottom 12% of the GFP expressing cells respectively. **b.** Fluorescence comparison between recovered High GFP and Low GFP sorted cells. The High GFP and Low GFP population were sorted out and recovered on SC-Leu plates. By photographing with the Bio-Rad Gel Doc system, the fluorescence of the recovered yeast colonies was recorded and compared. **c.** Fluorescence quantification of recovered GFP-sorted cells by flow cytometry. The average fluorescence intensity (Geometric Mean) of three individual colonies from the High GFP population is higher than that of those from the Low GFP population, while the geometric means of the three colonies from High GFP population are very different from each other. SCR H: SCRaMbLEed strain with higher fluorescence; SCR L: SCRaMbLEed strain with lower fluorescence.

Some recovered colonies were picked individually for further quantification of the fluorescence intensity. The result shows that colonies from the High GFP population have much higher fluorescence than the un-SCRaMbLEed strain, the SCRaMbLEed strain with GFP and also the Low GFP population (Figure 4.5c). Also, it is observed that the expression heterogeneity of the High GFP population is more obvious than

the Low GFP expression population. Therefore, the next step was using the sorted High GFP population for continuous SCRaMbLE and sorting to see whether after rounds of consecutive SCRaMbLE and sorting, the fluorescence intensity will be increased for the whole cell population. The experiment will be carried out by future lab members and once the final higher production strains are screened out, they will be sent for whole genome sequencing for genotype analysis.

4.2.1.3 Fluorescence distribution of fluorescent metabolites

Since the final aim of SCRaMbLEing the synthetic yeast is to optimize the yeast chassis for heterologous pathway expression, an automated screening method for target chemical product is also worth exploring. Taking the β-carotene pathway as an example, it was reported in a previous study that the carotenoid-expressing yeast showed fluorescence and the excitation/emission wavelength were similar to those of GFP (An, Bielich, Auerbach, & Johnson, 1991). Based on that, the fluorescence distribution of carotenoid producing yeast was tested with fluorescence microscopy and flow cytometry. The microscopy result shows that there are fluorescent spots distributed unevenly in the carotenoid-producing cells under the YFP and FITC channel (Figure 4.6a). The formation of these particles could be because of the aggregation of the hydrophobic carotenoids in the cellular environment. Next, the fluorescence distribution of different carotenoid production strains, which were generated with the *in vitro* DNA recombinase toolkit that will be described in Chapter 5, were compared. From LWy068 to LWy076, the degree of orange color gradually increased (see Figure 5.16b). The fluorescence distribution of these strains shows that fluorescence peaks are different between the non-carotenoid producing BY4741 and the most orange strain LWy076. However, there is a big overlap of the whole distribution of the two strains; for the rest of the strains, the distribution is between BY4741 and LWy076 (Figure 4.6b). With the combination of the microscopy and the flow cytometry results, it seems that the fluorescence of carotenoids might not be distinctive enough from the wild type and the heterogeneity could make it difficult to effectively screen out hyperproduction strains from the cell mixture. In addition,

FACS cannot precisely measure the composition and the quantity of the carotenoids. Therefore, direct selection on metabolites by FACS was not investigated further.

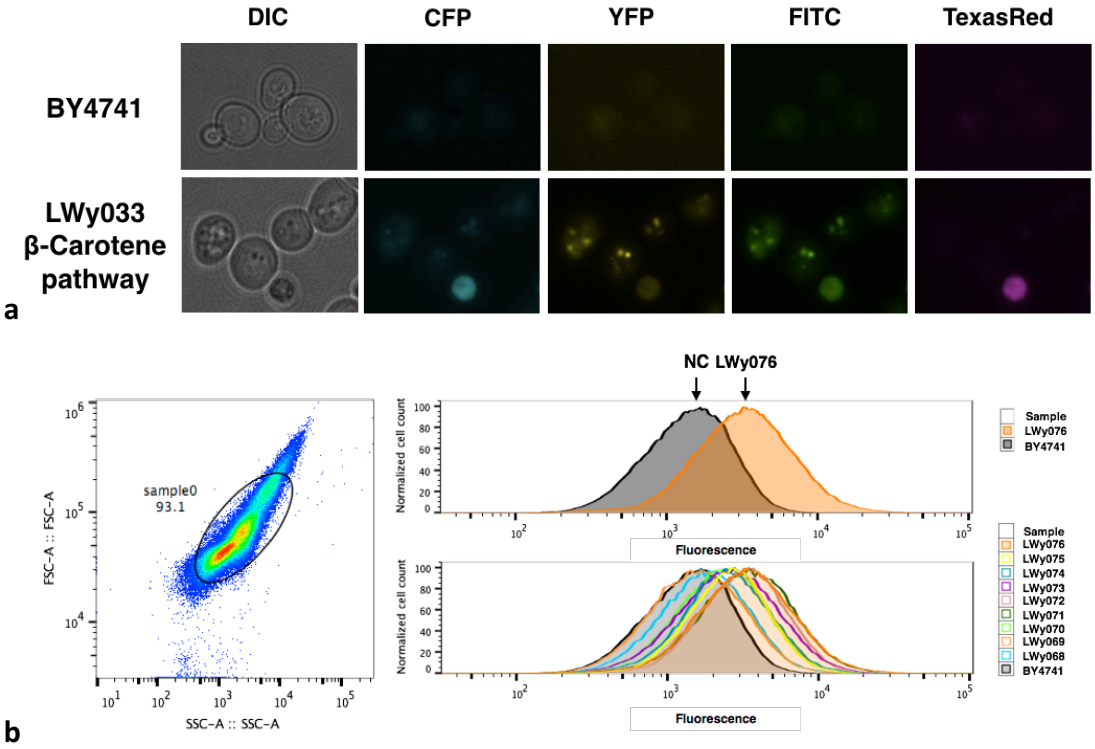


Figure 4.6|Fluorescence measurement of carotenoids in yeast. a. Fluorescence microscopy of carotenoid producing yeast. Fluorescent spots can be observed in the YFP and FITC channel. Wild type strain BY4741 was used as negative control. **b.** Fluorescence distribution of various carotenoid-producing yeast cells by flow cytometry. The fluorescence peak of the wild type control strain and the most orange carotenoid producing strain are separated but there is a big overlap of the distributions, shown in the upper panel. Nine more yeast strains with different carotenoid production levels were also tested with flow cytometry and there are overlaps of fluorescence distribution as well among all these strains, shown in the lower panel.

4.2.2 Screen and verification of SCRaMbLE-in strains

4.2.2.1 5-FOA counter selection

While preparing the automated screening method, the function of the SCRaMbLE-in device was characterized at the same time. Based on the universal SCRaMbLE-in acceptor, five SCRaMbLE-in devices were constructed: β-carotene, violacein, P_{TDH3}-GFP, P_{ACT1}-GFP and $P_{YGR270W}$-GFP. The genetic compositions of the β-carotene and

violacein pathways are as described in chapter 3 (Table 3.4). As described in the SCRaMbLE-in workflow, all SCRaMbLE-in devices were first transformed into the yeast strain with SynII chromosome LWy095. Then one strong Cre expression device, P_{TDH3}-CreEBD, one weak Cre expression device, P_{SCW11}-CreEBD and a control empty plasmid pRS413 were transformed afterwards. As an initial test, strains with three GFP SCRaMbLE-in devices and the different Cre expression devices were induced by 1μM estradiol for 12h and 24h. After the induction, the cultures were serially diluted and spotted onto SC-Leu and SC-Leu 5-FOA plates. The result shows that strains with the GFP SCRaMbLE-in devices and the pRS413 vector grew well on SC-Leu while no survivor was observed on the SC-Leu 5-FOA plate for both 12h and 24h induction; strains with the GFP SCRaMbLE-in devices and P_{TDH3}-CreEBD plasmid grew well on SC-Leu plate, while 3 out of 6 isolates (isolate 3,4,5) survived on the SC-Leu 5-FOA plate after 12h induction and 5 out of 6 isolates (isolate 1, 3-6) survived on the 5-FOA plate after 24h induction; strains with P_{SCW11}-CreEBD plasmid grew well on SC-Leu plate, while 1 out of 6 isolates (isolate 10) grew on 5-FOA plate after 12h induction and 5 out of 6 isolates (isolate 7-9, 11,12) survived after 24h induction (Figure 4.7a). From the spotting assay, we can see that without the presence of Cre the strains cannot survive on the 5-FOA plate, demonstrating the effectiveness of the counter-selection; the efficiency of SCRaMbLE-in is better under 24h induction than 12h induction; SCRaMbLE-in process driven by P_{TDH3}-CreEBD is slightly faster than P_{SCW11}-CreEBD but also has a stronger negative effect on the fitness of the synthetic strain; the growth status of the strains were not synchronized which indicates that the SCRaMbLE-in process happens randomly during the activation of recombination. Therefore, for the pathway SCRaMbLE-in experiment P_{SCW11}-CreEBD was used considering the burden of the heterologous expression of the pathway and P_{TDH3}-CreEBD's effect on the fitness of the synthetic yeast. The induction time also increased to 24h and 48h for the spotting assay. The result shows that the number of survivors on the 5-FOA plate after 24h induction was similar to that of the GFP SCRaMbLE-in; the number of survivors on the 5-FOA plate after 48h induction was similar to that on SC-Leu plate, meaning that the potential SCRaMbLE-in strain had been predominant in the cell

population after 48h induction and cell growth (Figure 4.7b). At the same time, the induced cells were diluted properly and plated to SC-Leu 5-FOA to get single colonies for PCR verification. The plating result shows that the colony size and color were diversified. Though they were not extremely diversified in this round of induction it is still worth doing deeper phenotype and genotype analysis for the potential SCRaMbLE-in strains (Figure 4.7c).

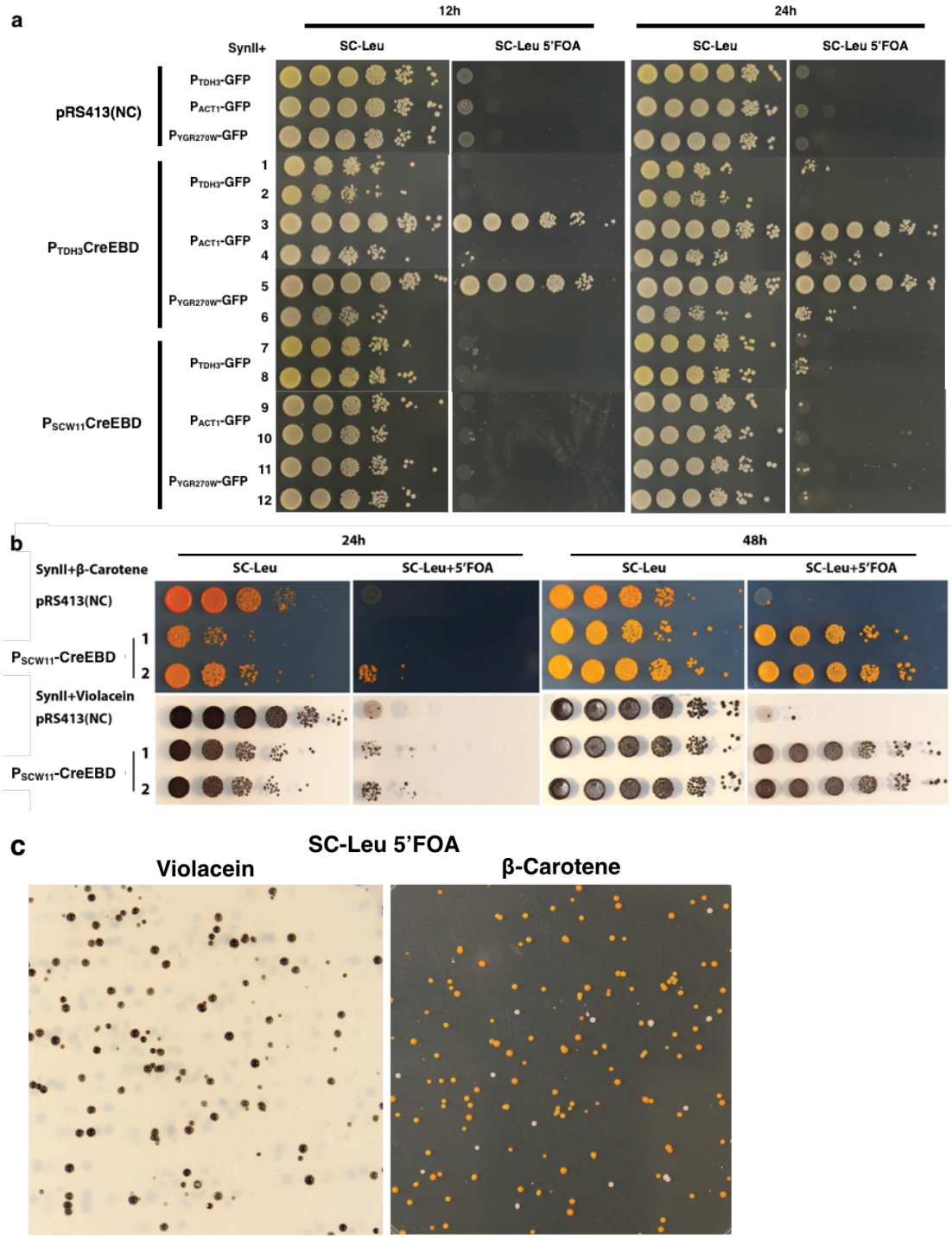


Figure 4.7|Spotting and plating assay of 5-FOA counter selection. a. Spotting assay result for GFP SCRaMbLE-in. Cre expression device P_{TDH3}-CreEBD and P_{SCW11}-CreEBD were used for driving the SCRaMbLE-in process, pRS413 was used as a negative control. With 12h induction, 3 out of 6 strains with P_{TDH3}-Cre have survivors on the 5-FOA plate and 1 out of 6 strains with P_{SCW11}-Cre have survivors on the 5-FOA plate. With 24h induction, for both strains with P_{TDH3}-Cre and P_{SCW11}-Cre, 5 out of 6 strains have survivors on 5-FOA plate. For the negative control, there was no growth on the 5-FOA plate. **b.** Spotting assay result for pathway SCRaMbLE-in. Cre expression device P_{SCW11}-Cre was used for pathway SCRaMbLE-in. With 24h and 48h induction, cells grew on the 5-FOA plate. A few colonies grew on the 5-FOA plate for the negative control, which could be due to the long sequence of the pathway and possible homologous recombination. **c.** SC-Leu 5-FOA plating result for pathway SCRaMbLE-in. The plating shows diversified colony sizes and colors for synthetic yeast with violacein and β-carotene pathways after SCRaMbLE-in induction.

4.2.2.2 Screening verification by colony PCR

After the plating, single colonies from the 5-FOA plate were picked for both replica plating and colony PCR. Taking the pathway SCRaMbLE-in candidates as an example, five colonies from each pathway after counter selection were re-spotted to SC-Leu plates and then replica plated to SC-Ura plates to see whether there were any false positive strains still expressing *URA3*. The result shows that the counter-selection is effective to exclude *URA3*-expressing strains (Figure 4.8a). Besides replica plating, the spotting assay also shows that the colors of the samples of either violacein or β-carotene are varying from the original control sample (Figure 4.8a). To ensure the absence of the *URA3* gene, a pair of primers were designed to amplify the whole *URA3* expression cassette. The colony PCR result shows that *URA3* was absent in all of the selected colonies (Figure 4.8b). As an additional test, another pair of primers were used to amplify specifically the CEN/ARS region of the SCRaMbLE-in device to see whether there was any plasmid left in the cell. The PCR result shows that all β-carotene pathway candidates have no CEN/ARS amplification while 2 out of 5 of violacein pathway have signal (Figure 4.8b). Since 10 samples might not be enough to reflect the whole population, 46 more samples of all GFP, violacein and β-carotene SCRaMbLE-in strains were tested with the *URA3* and CEN/ARS tag. The result shows

that *URA3* was negative for all samples, while the CEN/ARS signal was negative for all samples from GFP and β-carotene strains but was positive for several samples from the violacein strains, which could be due to the larger size of the violacein gene cluster (Figure 4.8c). To this end, preliminary tests of all the selected SCRaMbLE-in candidates were done and those with negative signals for PCR tags were forwarded for whole genome sequencing to confirm the location of integration and other genome variations caused by SCRaMbLE.

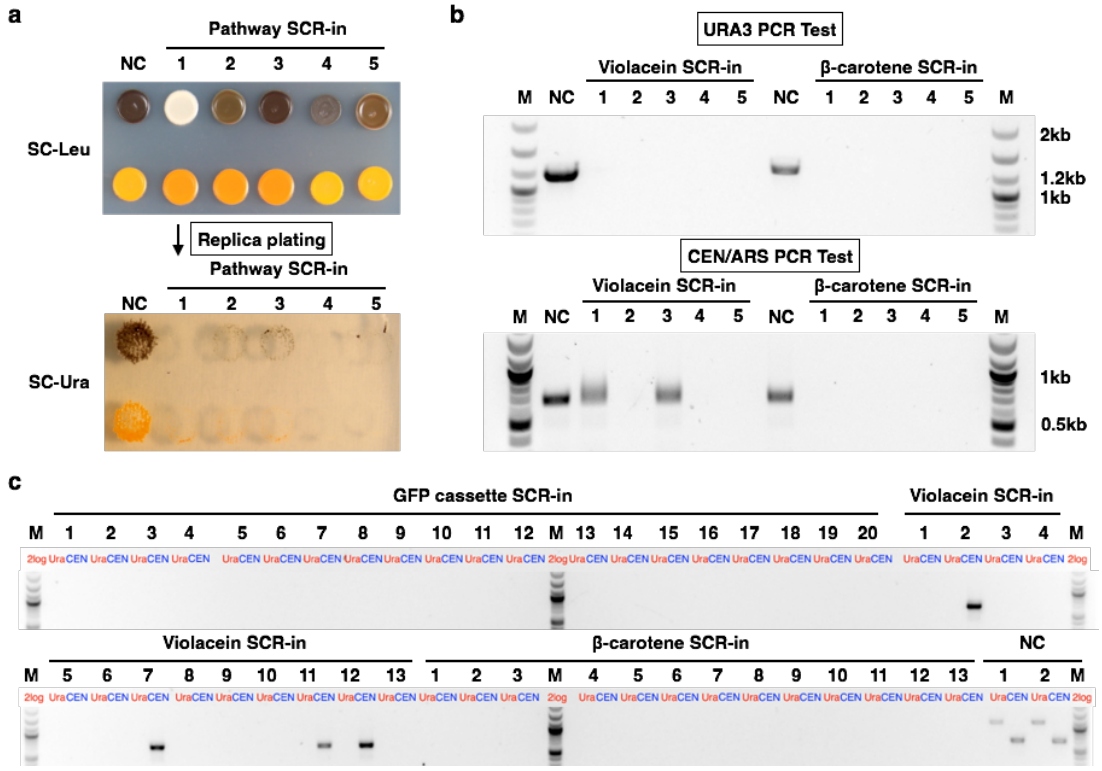


Figure 4.8|Preliminary verification of SCRaMbLE-in. a. Spotting and replica plating assay of 10 samples from pathway SCRaMbLE-in screening. **b.** *URA3* and CEN/ARS PCR test for the 10 pathway SCRaMbLE-in samples. **c.** *URA3* and CEN/ARS PCR test for 46 samples from GFP and pathway SCRaMbLE-in screening. *URA3* tag: 1.2kb; CEN/ARS tag: 0.7kb. NC: SynII + GFP/violacein/β-carotene SCRaMbLE-in.

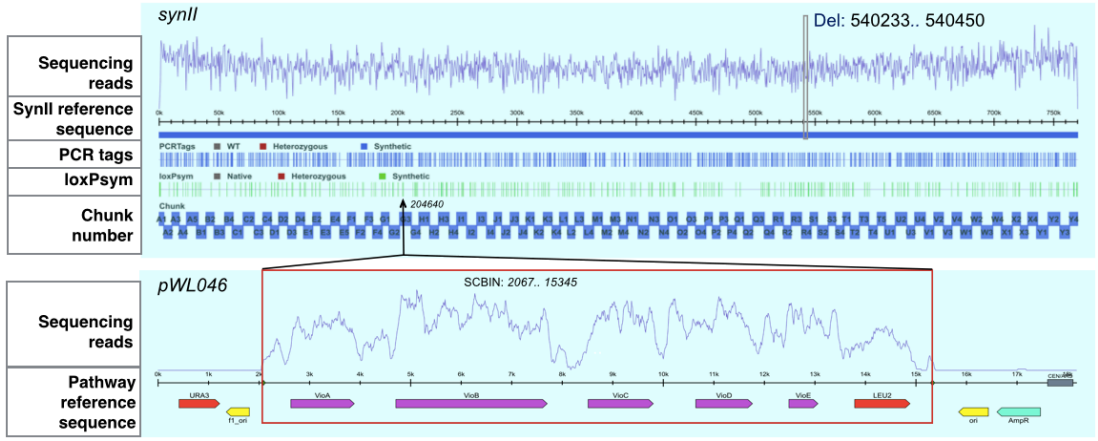
4.2.3 Pathway integration confirmation and pathway production comparison

4.2.3.1 Next generation sequencing of SCRaMbLE-in strains

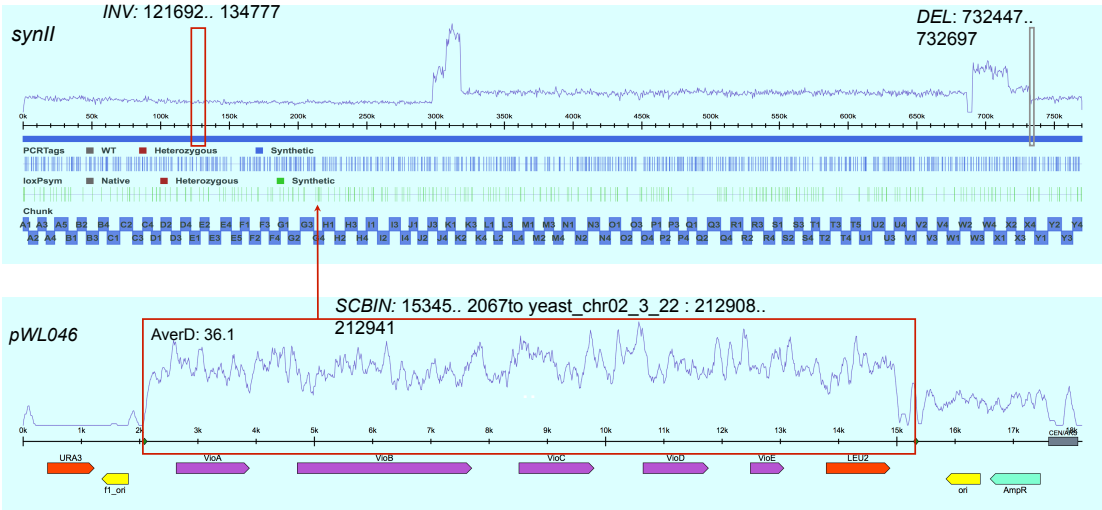
Strains which passed the initial PCR tag verification were sent for whole genome sequencing in Beijing Genomics Institute (BGI). The sequencing results show that the whole violacein and β-carotene pathway can be integrated to the synthetic chromosome II (GenBank accession: CP013608.1) at the loxPsym sites at different

locations in the chromosome. The sequencing results of six strains, LWy152, LWy238, LWy215, LWy252, LWy253 and LWy250, are displayed below (Figure 4.9). Five out of six strains were integrated to a loxPsym site on SynII but one was integrated to the *HIS3* locus on Chromosome XV. The locations of integration on SynII are 204640, 212908, 281280, 298247 and 19952 for LWy152, LWy238, LWy215, LWy252 and LWy253 respectively. The centromere location of SynII is 215886 to 215947, which is very interesting since 4 out of 5 SCRaMbLE-in locations are close to the centromere and the other one is closer to the telomere of the chromosome. Though the sample number is not yet enough to draw any conclusions about the relationship between location and frequency of the integration, one explanation could be that those with darker color were selected and sent for NGS and these locations could enhance pathway expression and production. The chemical analysis of these strains will be introduced in the next section. Besides pathway integration, other genome variations including deletion, inversion and duplication were also observed. Among these strains, LWy238 has the most dramatic and complicated genome rearrangement (Figure 4.9b). Besides simple inversion and deletion, the genome of LWy238 increases around 394Kb in length, including 2 to 7 copies of certain fragments between regions from 297781 to 720928 (Figure 4.9c). Since the duplicated area is large and the variation pattern is very complicated, the DNA order can only be narrowed down to 72 possible combinations; one order is displayed in Figure 4.9c. The selection of SCRaMbLE-in strains added pressure and complicated genome rearrangements were generated with this strategy for pathway diversification. However, besides the loxPsym site integration, off-target integration to Chromosome XV was also observed (Figure 4.9g). Instead of a loxPsym site on SynII, the pathway is inserted to position 721628 on the mutated *HIS3* region of ChrXV. Considering that the Cre expression cassette is on a *HIS3* marked plasmid, the off-target integration could be related to homologous recombination indirectly or it could be that there are pseudo-loxP sites on ChrXV to accept the pathway. The whole genome sequencing result confirmed the integration of the pathway into SynII but also pointed out some noise outside the planned integration.

a. Violacein SCRaMbLE-in - LWy152

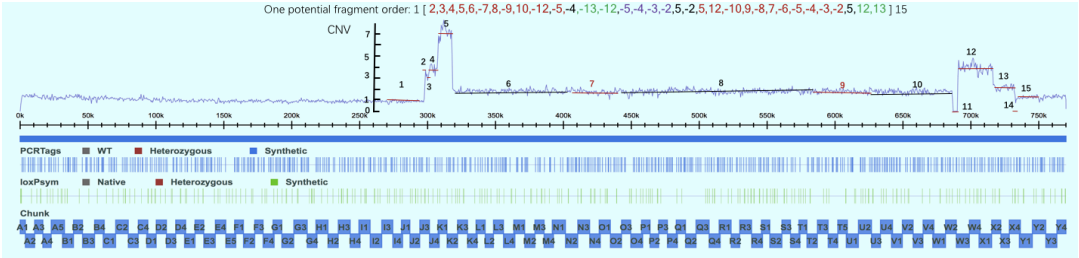


b. Violacein SCRaMbLE-in – LWy238

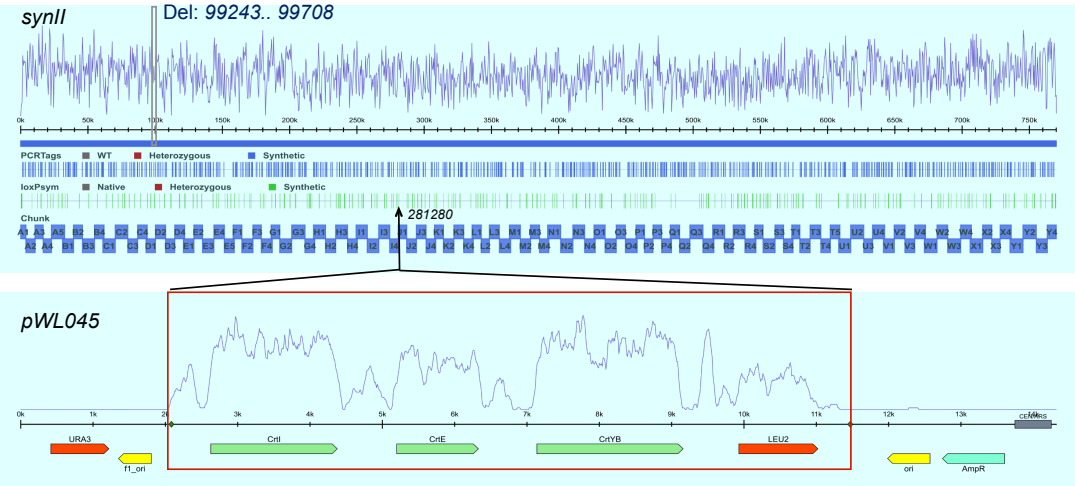


c. LWy238, large fragment duplication and inversion

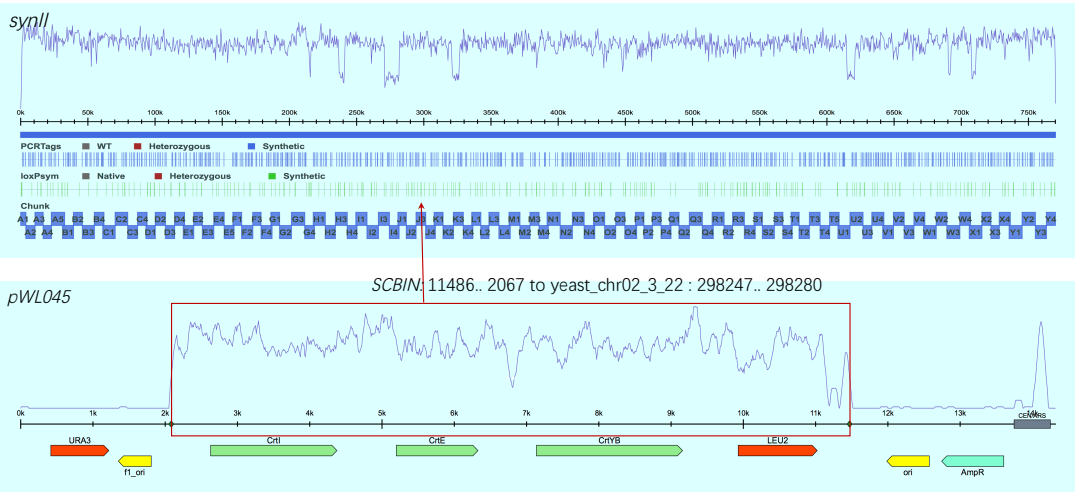
FragID	Region	Depth	CN
1	1 .. 297748	42.3	1
2	297781 .. 298247	175.4	4
3	298280 .. 300615	127.4	3
4	300648 .. 307359	162.5	4
5	307392 .. 317884	299.8	7
6	317917 .. 403656	76.2	2
12	690074 .. 716037	164.1	4
13	716070 .. 720928	86.8	2



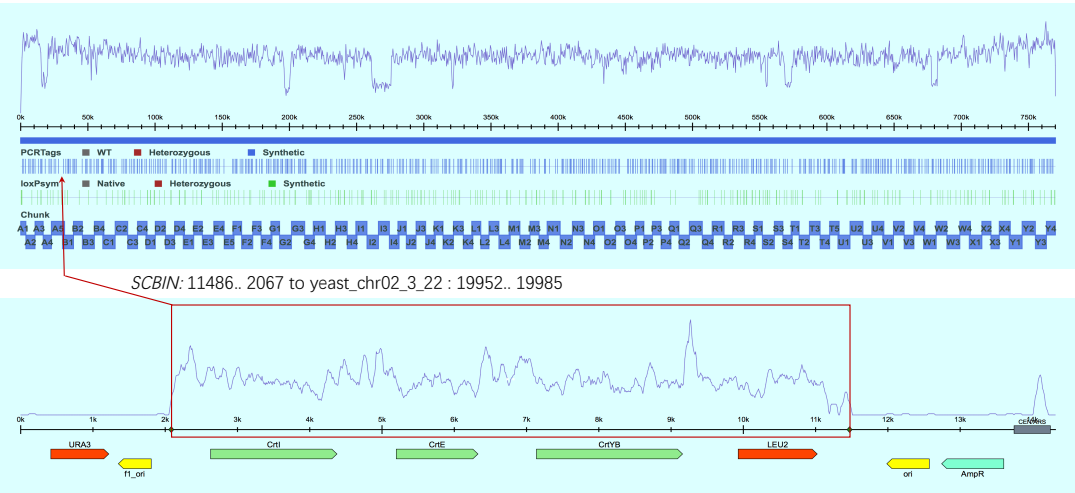
d. β-carotene SCRaMbLE-in – LWy215



e. β-carotene SCRaMbLE-in – LWy252



f. β-carotene SCRaMbLE-in – LWy253



g. Violacein off-target integration in ChrXV

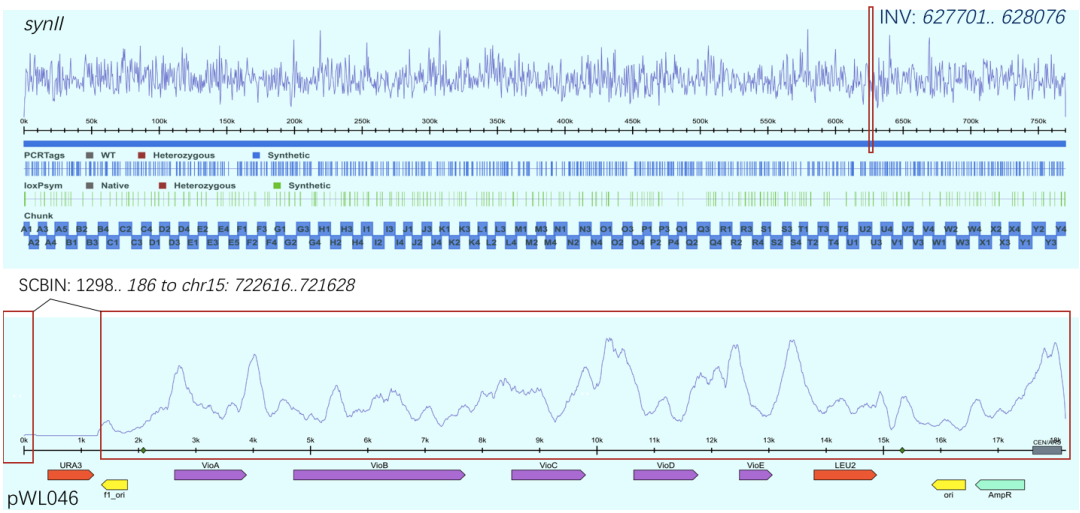
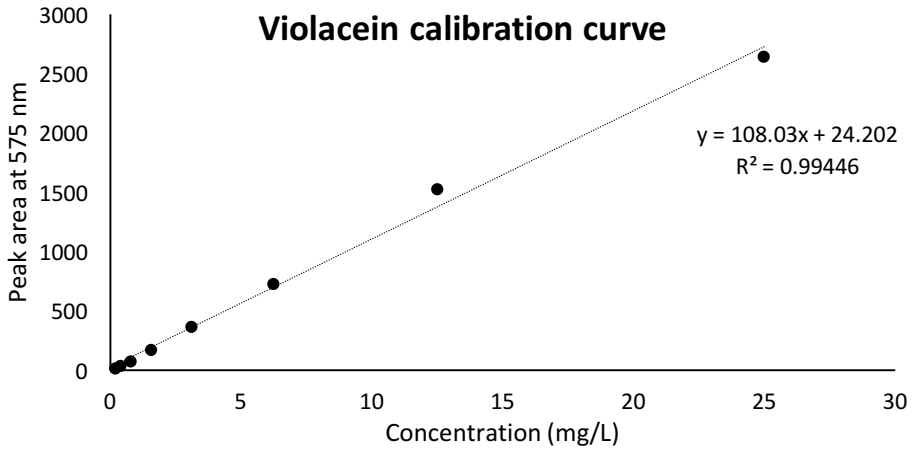


Figure 4.9|Whole genome sequencing of the SCRaMbLE-in strains. a. LWy152, violacein pathway integrated to the loxPsym site at position 204640, with 217bp deletion from position 540233 to 540450. **b.** LWy238, violacein pathway integrated to the loxPsym site at position 212908, with a 13kb inversion from 121692 to 134777, a 250bp deletion from 732447 to 732697 and a complex combined inversion and duplication from 297781 to 720928. **c.** Copy number variation (CNV) and potential order of the complex genome rearrangement of LWy238. According to the average depth of sequencing, the copy number of fragments varies from 2 to 7 and the total additional sequence is around 394kb. One potential order of the rearranged piece is listed. **d.** LWy215, β-carotene pathway integrated to the loxPsym site at position 281280, with a 465bp deletion from 99243 to 99708. **e.** LWy252, β-carotene pathway integrated to the loxPsym site at position 298247 with no extra variation. **f.**LWy253, β-carotene pathway integrated to the loxPsym site at position 19952 with no extra variation. **g.** LWy240, violacein pathway off-target integration to *HIS3* locus on Chromosome XV. For LWy240, there is an inversion from 627701 to 628076 on *SynII* but the violacein pathway is not inserted to SynII but to ChrXV at 721628 instead.

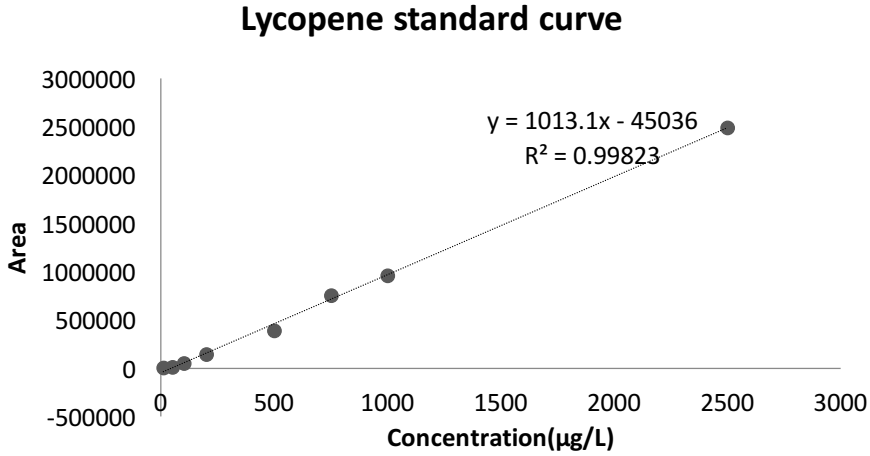
4.2.3.2 Production variation of violacein and β-carotene after SCRaMbLE-in

After the whole genome sequencing was done for the selected SCRaMbLE-in variants, the chemical composition and production of the SCRaMbLE-in variants were also characterized and quantified. To quantify the production of violacein and the two major carotenoids, β-carotene and lycopene, commercial standards of the three products were purchased. HPLC was used for violacein quantification and LCMS was used for β-carotene and lycopene quantification. With proper dilution, standard

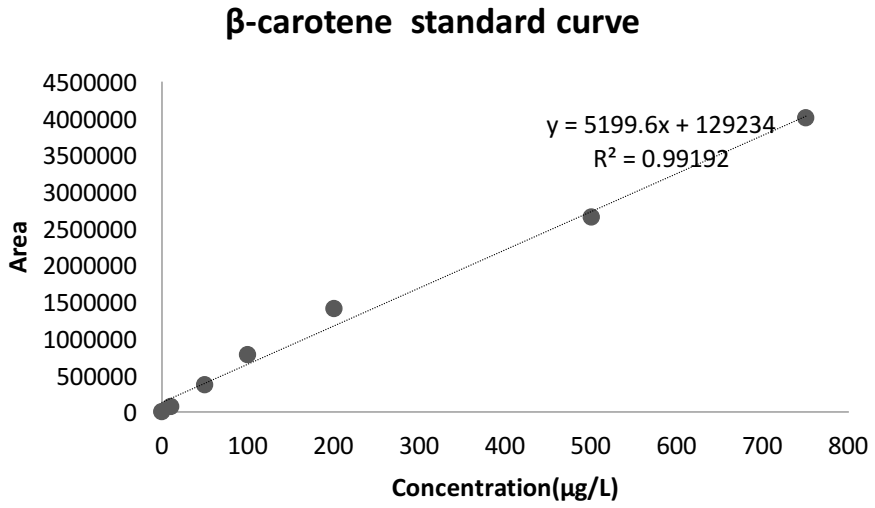
curves were made for the three chemicals (Figure 4.10). Chemical extraction and quantification for the five NGS confirmed SCRaMbLE-in strains were carried out with the same HPLC or LCMS method described in section 2.13. Strain LWy134 and LWy137 with pathways inserted at the HO-locus were used as control. The result shows that HO-inserted violacein strain LWy137 has a yield of around 1mg/L while in the SCRaMbLE-in strains LWy152 and LWy238 the yields are around 5mg/L, almost five fold that of the control strain (Figure 4.11a). For the β-carotene pathway, the total carotenoid quantity for HO integrated strain LWy212 is around 1000 μg/L ; the carotenoid quantities of the two SCRaMbLE-in strains LWy215 and LWy252 are similar, but LWy215 is slightly lower than the control while LWy252 is higher than the control; the total carotenoid of LWy253 is around 500 μg/L which is half that of the control strain (Figure 4.11b). However, there are differences in the quantity or ratio of β-carotene and lycopene among the SCRaMbLE-in strains. For LWy215, the quantity of both β-carotene and lycopene is reduced; for LWy252, the quantity of β-carotene is increased to twice of that of the control strain while the quantity of lycopene is slightly reduced, indicating higher transformation efficiency from lycopene to β-carotene; for LWy253, though the total carotenoid quantity is reduced to half, the yield of β-carotene is similar to that of the control strain, which also indicates a higher transformation rate of lycopene to β-carotene. Both the genotype and phenotype were thus characterized for the SCRaMbLE-in strains and proven to be diversified by the activation of Cre recombinase. The effects of genome integration and genome rearrangement by SCRaMbLE were different for the two demonstration pathways and the SCRaMbLE-in device can be expanded for integrating more pathways into more synthesized chromosomes in future.



a.



b.



c.

Figure 4.10| Calibration curve for violacein and carotenoids. a. HPLC calibration curve for violacein. **b.** LCMS calibration curve for lycopene **c.** LCMS calibration curve for β-carotene.

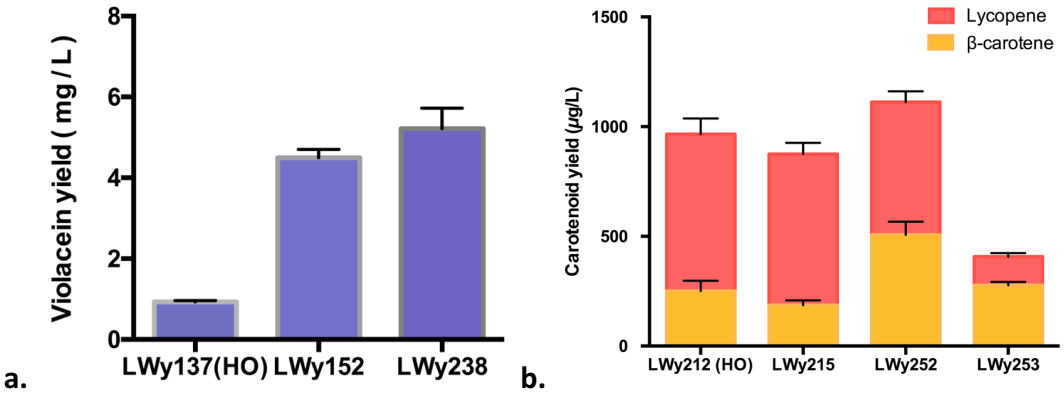


Figure 4.11 | Violacein and carotenoid quantification of SCRaMbLE-in strains. a. HPLC quantification of violacein SCRaMbLE-in strains. The violacein yield of two SCRaMbLE-in strains, LWy152 and LWy238, was around 5mg/L, which is 5-fold higher than that of the HO-integrated control strain LWy137. **b.** LCMS quantification of β-carotene SCRaMbLE-in strains. SCRaMbLE-in strain LWy252 has the highest yield of both total carotenoids and β-carotene. LWy215 has slightly lower production of total carotenoids and β-carotene than HO-integrated control strain. LWy253 has half the production of total carotenoids but similar β-carotene production as the HO integrated control strain.

4.2.4 Continuous SCRaMbLE for violacein SCRaMbLE-in strains

Since the genome rearrangements generated by SCRaMbLE are random and the possibility of genome variation is nearly infinite, one round of SCRaMbLE-in activation can diversify the production of a pathway but may not be enough to optimize the chassis just in one-time induction. Therefore, after the first step of SCRaMbLE-in, the strains with a successfully integrated pathway can be sent for continuous SCRaMbLE to generate more genome variants and to screen for mutants with even higher production. Strains with lower production are also interesting to look into since we can understand more how the host genome context can affect heterologous expression of a pathway. Considering that the chemical production of the violacein pathway can be roughly estimated from the color of the colony, violacein-integrated SynII strains were chosen for a continuous SCRaMbLE experiment. Three NGS-verified SCRaMbLE-in strains LWy152, LWy238 and LWy239, were re-transformed with P_{SCW11}CreEBD for continuous SCRaMbLE. After 24h of β-estradiol induction, the cells

were plated to SC-Leu plates to get single colonies for color comparison. On the plate of SCRaMbLEed LWy152, LWy238 and LWy239, colonies were found with both darker (green arrow) and lighter (red arrow) color, which indicates increased and decreased pathway expression (Figure 4.12a). The samples were then standardized to OD_{600}=1.0 at log phase with measurement by spectrophotometer and spotted to SC-Leu to compare the color difference more clearly (Figure 4.12b).

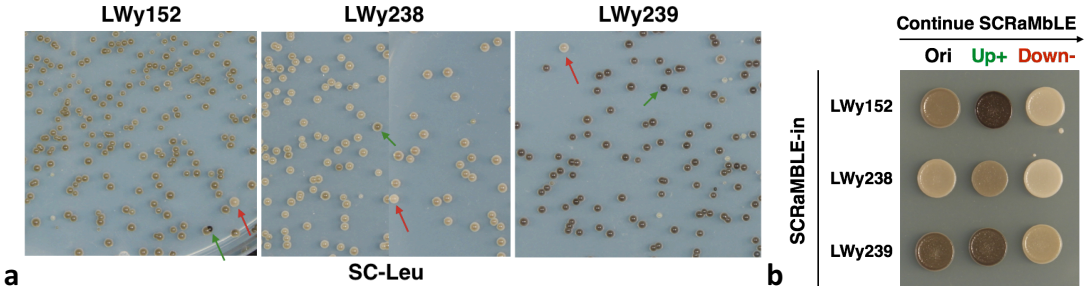


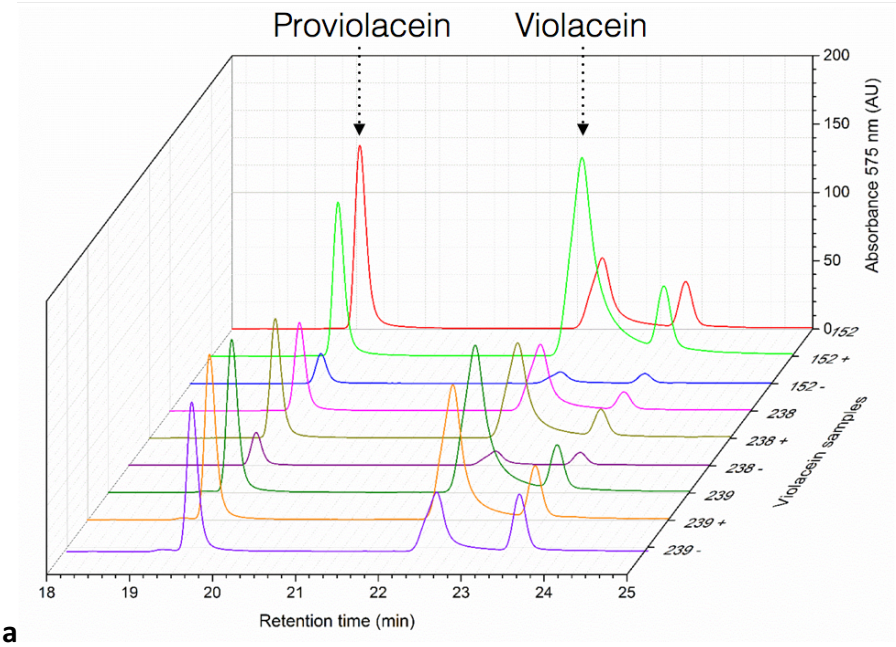
Figure 4.12|Continuous SCRaMbLE screening of violacein pathway. a. Cell plating screening of SCRaMbLEed strains. Darker colonies were pointed with green arrow and lighter colonies were pointed with red arrow. **b.** Spotting of selected continuous SCRaMbLEed strains. The spotting assay confirmed the color variation of the candidate continuous SCRaMbLEed colonies.

4.2.5 Phenotype and genotype characterization of continuous SCRaMbLEed strains

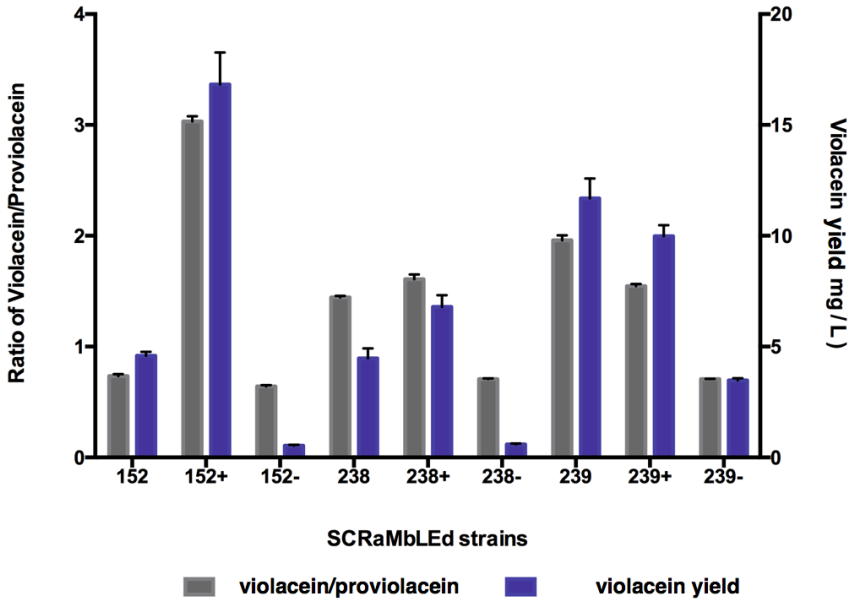
4.2.5.1 Production variation of violacein with continuous SCRaMbLE

After the initial plate screening by color variation, the composition of violacein and its ratio with the intermediate proviolacein were quantified for the continuous SCRaMbLE strains. The HPLC result shows that the peak areas of violacein and proviolacein were diversified among the SCRaMbLEed strains (Figure 4.13a). With the calibration curve of violacein, the yield of violacein and the conversion efficiency from Proviolacein to violacein were quantified. LWy152 upregulated strain LWy152+ has the highest violacein yield of 16.8mg/L, with more than 3-fold that of LWy152 and nearly 17-fold that of the HO integrated control strain; the yield of LWy238+ is 1.5-fold that of the LWy238 strain; however, the yield of LWy239+ is not higher than that of LWy239 and is not consistent with the color change judged by observation, which illustrates that precise chemical quantification is necessary and a more reliable way

of characterization. For the downregulated strains, LWy152- and LWy238-, their violacein production was reduced to around 0.5 mg/L, which is almost 1/10 of the production of either LWy152 or LWy238; the yield of LWy239- is reduced to 3.5mg/L but is not as dramatically reduced as LWy152- and LWy238-. As to the ratio of violacein to proviolacein, LWy152, LWy152-, LWy238- are similar, around 0.7; LWy238, LWy238+, LWy239 and LWy239+ are similar, from 1.4 to 1.9; LWy152+ still has the highest of 3.0 (Figure 3.13b).



a



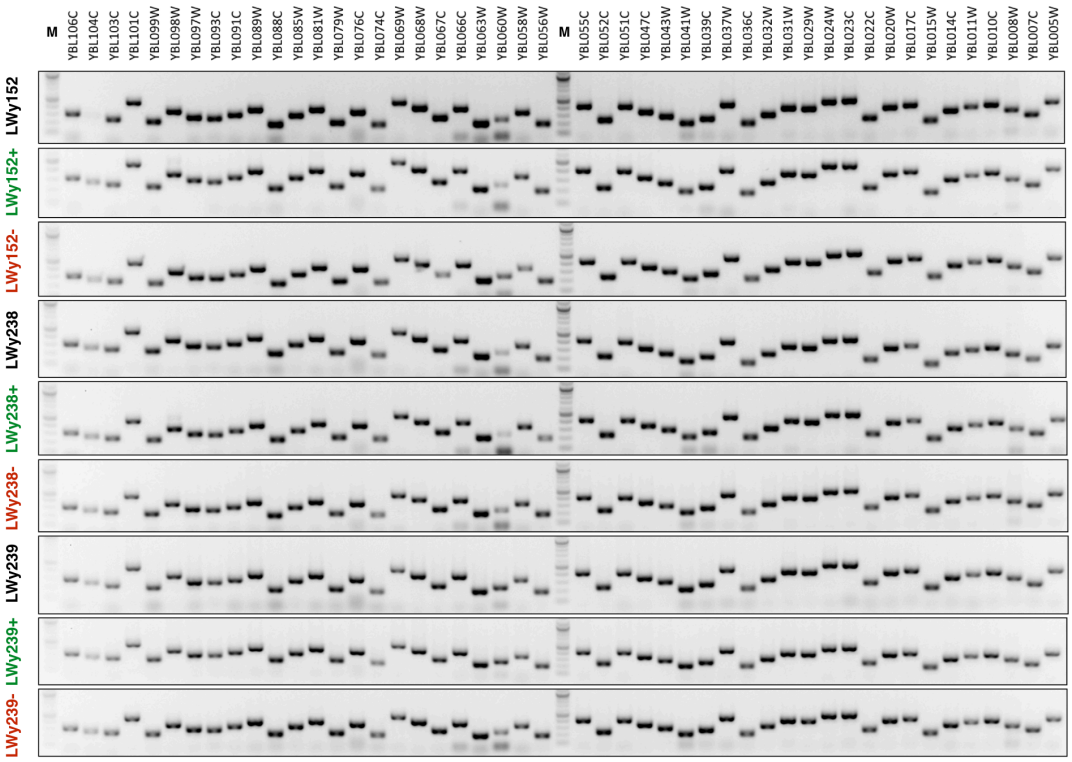
b

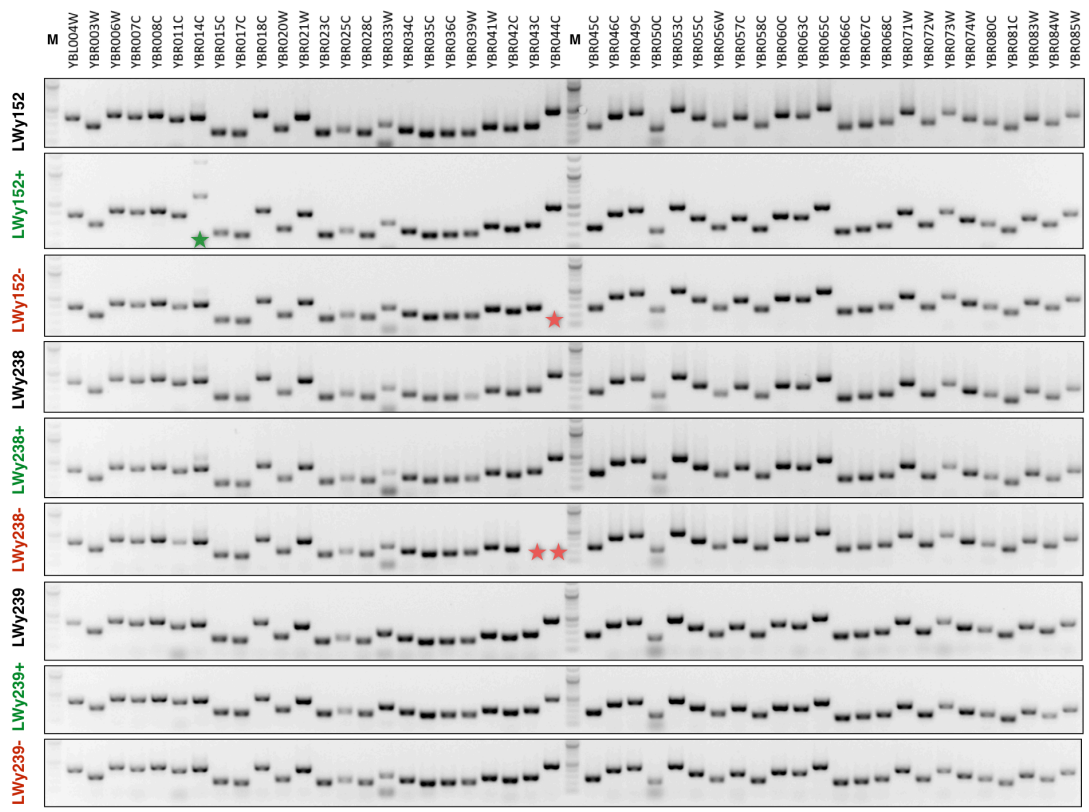
Figure 4.13|HPLC quantification of violacein continuous SCRaMbLEed strains. a. Chromatograms of multiple continuous SCRaMbLEed strains with HPLC. The graph was plotted for multiple strains for easier comparison of the chemical compositions between the mutants. The absorbance was measured at 575mm, which is the absorption peak for violacein. **b.** Quantification of violacein/proviolacein ratio and violacein production. The violacein/proviolacein ratio is shown in grey and the violacein production is shown in purple. The conversion efficiency of proviolacein to violacein and the violacein yield were all confirmed to be diversified by continuous SCRaMbLE. Strain LWy152+ was upregulated from the LWy152 strain after conti-SCRaMbLE and has the highest violacein/proviolacein ratio and also the highest violacein yield of 16.8mg/L. Biological replicates (n=3) were performed for the HPLC quantification.

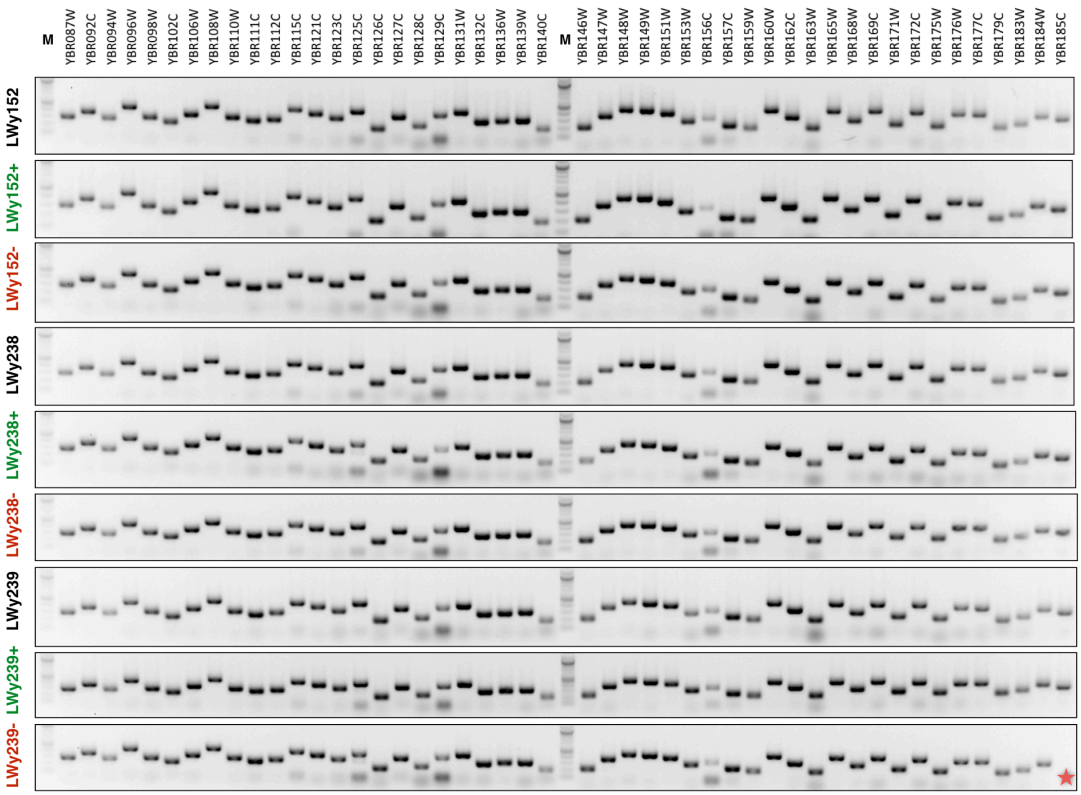
4.2.5.2 SynII PCR tag assay

After the chemical quantification of the SCRaMbLEed strains was performed, we wanted to dig deeper into the genotype variation of the strains. Instead of sending the strains directly to next generation sequencing, a pre-test of fragment deletion by synthetic PCR tags was done. Since one of the design principles of the yeast synthetic chromosome is putting PCRtagging watermarks in the genome to distinguish the synthetic sequence from the wild type sequence, a library of PCR primers were designed to monitor the construction process of the synthetic chromosomes (Richardson et al., 2017). From the SynII primer library, 192 primer pairs were chosen by Dr. Abramczyk in the Cai laboratory to amplify fragments between loxPsym sites with coding sequence. Therefore, if a signal for a primer pair is missing, it indicates CDS deletion between the two loxPsym sites. The SynII PCR tag assay on the 9 SCRaMbLEed strains shows that LWy152+ has 1 missing tag, LWy152- has 1 missing tag, LWy238,238+,238-, LWy239, 239+, 239- all have one same missing tag, LWy238- has two extra missing tags and LWy239- has one extra missing tag (Figure 4.14). Since not all genome variations can lead to the change of phenotype, it is necessary to narrow down the genome rearrangements that lead to the phenotype variation. One possible solution could be finding shared genotype variations in strains with common phenotype variation. When looking at the deletion patterns of the mutant strains, most of them are independent from each other. Though LWy238,238+,238-, LWy239,

239+, 239- have the same deletion in YBR250W, it could be because the continuous SCRaMbLEed strains inherited the genome deletion from the source strain and the NGS sequencing results of LWy238 and LWy239 show that they have derivative relationships, making YBR250W a less special target to work on. On the other hand, LWy152- and LWy238- were SCRaMbLEed from different source strains but both have the same PCR tag deletion of YBR044C. Considering that the two strains were both down-regulated in violacein production to similar level (Figure 4.13b), further tests on the effect of YBR044C on the downregulated phenotype change were performed and will be described in next section. Besides tag deletion, there is also a size variation from one 400bp band to two bands with 600bp and 1.2Kb in LWy152+ at YBR014C. The phenomenon could indicate potential inversion or duplication events in the area, but only NGS confirmation will tell the exact rearrangement for further genotype and phenotype analysis.




M
YBL004W
YBR003W
YBR006W
YBR007C
YBR008C
YBR011C
YBR014C
YBR015C
YBR017C
YBR018C
YBR020W
YBR021W
YBR023C
YBR025C
YBR028C
YBR033W
YBR034C
YBR035C
YBR036C
YBR039W
YBR041W
YBR042C
YBR043C
YBR044C
M
YBR045C
YBR046C
YBR049C
YBR050C
YBR053C
YBR055C
YBR056W
YBR057C
YBR058C
YBR060C
YBR063C
YBR065C
YBR066C
YBR067C
YBR068C
YBR071W
YBR072W
YBR073W
YBR074W
YBR080C
YBR081C
YBR083W
YBR084W
YBR085W
LWy152
LWy152+
LWy152-
LWy238
LWy238+
LWy238-
LWy239
LWy239+
LWy239-


M
YBR087W
YBR092C
YBR094W
YBR096W
YBR098W
YBR102C
YBR106W
YBR108W
YBR110W
YBR111C
YBR112C
YBR115C
YBR121C
YBR123C
YBR125C
YBR126C
YBR127C
YBR128C
YBR129C
YBR131W
YBR132C
YBR136W
YBR139W
YBR140C
M
YBR146W
YBR147W
YBR148W
YBR149W
YBR151W
YBR153W
YBR156C
YBR157C
YBR159W
YBR160W
YBR162C
YBR163W
YBR165W
YBR168W
YBR169C
YBR171W
YBR172C
YBR175W
YBR176W
YBR177C
YBR179C
YBR183W
YBR184W
YBR185C
LWy152
LWy152+
LWy152-
LWy238
LWy238+
LWy238-
LWy239
LWy239+
LWy239-

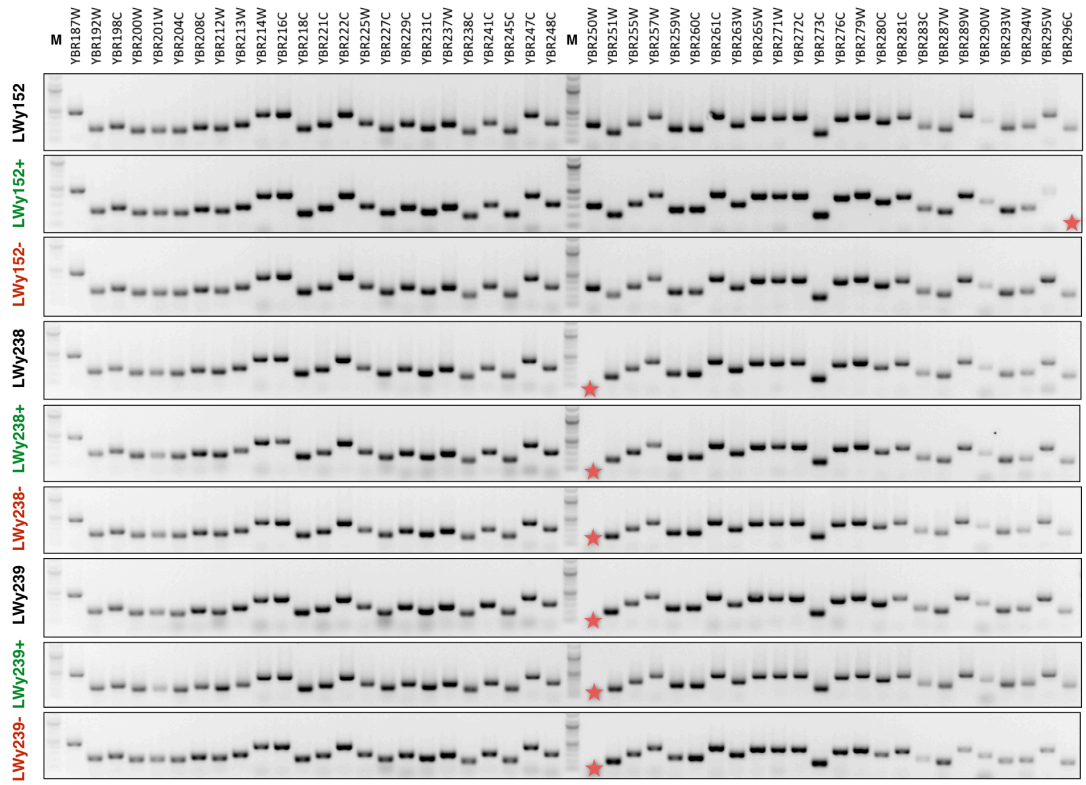


Figure 4.14|Synthetic PCR tag verification of the continuous SCRaMbLE strains. LWy152 has no missing tags; LWy152+ has a missing tag at YBR296C and a modified tag at YBR014C; LWy152- has a missing tag at YBR044C; LWy238,238+,238-,239, 239+,239- all have a missing tag at YBR250W; LWy238- has two extra missing tags at YBR043C and YBR044C; LWy239- has an extra missing tag at YBR185C. Red star indicates fragment deletion; green star indicates potential fragment duplication or other types of variation.

4.2.5.3 Gene complementation experiment for strains with common deleted tags

From the PCR tag assay, we knew that the down-regulated LWy152- and LWy238- strain have the same deleted tag YBR044C. To investigate whether it is due to the deletion of the YBR044C gene that the two strains have lower violacein production, a gene complementation experiment was done on the two strains. The YBR044C transcription unit was cloned to both the low-copy centromeric vector pRS413 and the high-copy episomal vector pRS423 for transformation back to the SCRaMbLEed strains. The purpose of putting high-copy cloned YBR044C is to see whether overproduction of the related deleted gene could improve the production of violacein. Together with a negative control vector pRS413, the two YBR044C

constructions were transformed to LWy152, 152+, 152-, 238 and 238- for color comparison. Two colonies from each transformation were picked and spotted with a standardized OD_{600} of 1.0 on the SC-His plate. The result shows that with the supplement of pRS413-YBR044C, LWy152- and LWy238- both recovered to the same color as the source strain LWy152 and LWy238, indicating that YBR044C contributes to the reduced violacein yield (Figure 4.15). However, the adding of pRS413-YBR044C to LWy152 and LWy152+ did not darken the color of the two strains; the overproduction of YBR044C did not improve the violacein yield but added burden to the cell (Figure 4.15). Nonetheless, the complementation experiment shows that the contribution of various genome rearrangements to a phenotype change by SCRaMbLE can be narrowed down through detecting shared deleted tags and can be attempted for testing other types of genome variation in future with the NGS data. The standard name of YBR044C is TCM62, which is a mitochondrial membrane protein and a putative chaperone involved in the assembly of the succinate dehydrogenase (SDH) complex (Dibrov, Fu, & Lemire, 1998). The deletion of TCM62 won't affect the viability of the cell but can interrupt the function of SDH and further affect the tricarboxylic acid cycle and the mitochondrial respiration chain in *S. cerevisiae.* As to how the deletion of YBR044C results in reduced violacein production, a potential explanation could be that the violacein synthesis process is NADPH-dependent; the deletion of TCM62 might affect the reducing power of the cell, indirectly affect the normal NADPH level and impede the transformation to violacein. It could also be that TCM62 has an unknown direct function on the pathway and this new phenotype of the tcm62 null mutant can also be added to SGD (Saccharomyces Genome Database) for future reference. Beside the shared deletion of YBR044C in LWy152- and LWy238-, YBR185C that is the sole deletion in LWy239-, encoded a membrane-associated mitochondrial ribosome receptor. The deletion of YBR185C can affect the process of mitochondrial translation, which probably further interrupts the normal respiratory chain and affect the synthesis of violacein. Though complementary experiments for YBR185C was not performed due to time limitation, it should be noticed the rule of mitochondrial function in violacein synthesis.

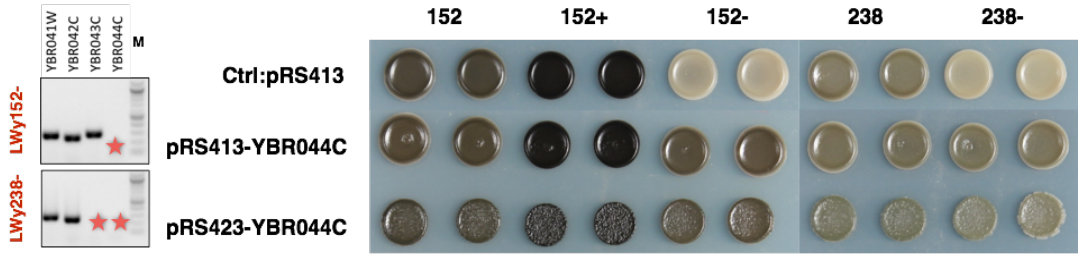


Figure 4.15|YBR044C complementation result. LWy152- and LWy238- have the same YBR044C deletion. LWy152- and LWy238- complemented with pRS413-YBR044C recovered the phenotype as the LWy152 and LWy238 strain respectively. Transformation of pRS413-YBR044C into LWy152, LWy152+ and LWy238 did not obviously improve the production of violacein. Overproduction of YBR044C on pRS423 vector did not obviously improve the production of violacein but affected cell fitness instead.

4.2.5.4 Whole genome sequencing of the continuous SCRaMbLEed strains

The PCR tag assay is a fast way of detecting fragment deletion but for fragment duplication, inversion or translocation, whole genome sequencing is still necessary for confirming all variations. The genome rearrangements for the source strains generated by SCRaMbLE-in and continuous SCRaMbLEed strains are summarized in Table 4.3. The results show that besides the deletion detected with the PCR tag assay, there are more complex genome rearrangements in the continuous SCRaMbLEed strains, including deletions of UTR regions, inversions and duplications. Since the loxPsym sites were all located outside CDS, the inversion of fragments normally only switches the UTR regions of two genes, which could have minor impact on the expression level of the genes. The deletion of 3'UTR regions will also affect the expression level of genes. The duplication of a CDS will affect its expression level while the deletion of a CDS will lead to total function loss of the gene and might trigger consequent changes in metabolic fluxes. The additional genome variations of the continuous SCRaMbLEed strains provide information on candidates that can be related to the change of phenotype and further experiments such as gene complementation or gene overproduction can be designed to explain the mechanisms of up- or down- regulation of the pathway. This work can be carried on in future for deeper studies.

Table 4.3 NGS summary of the continuous SCRaMbLEed strains

Sample_ID	Chr/Plasmid ID	1st_loxP start	2nd_loxP end	Variation type
LWy152	SynII 1.3	204640	204673	INS
	pWL046	2067	15345	INS_fragment
	SynII 1.3	540233	540450	DEL
LWy256 (LWy152+)	SynII 1.3	204640	204673	INS
	pWL046	2067	15345	INS_fragment
	SynII 1.3	**204115**	**204673**	**INV**
	SynII 1.3	**215628**	**216620**	**INV**
	SynII 1.3	**236489**	**237846**	**DEL**
	SynII 1.3	**240777**	**241151**	**INV**
	SynII 1.3	540233	540450	DEL
	SynII 1.3	**619127**	**619127**	**INV**
LWy257 (LWy152-)	SynII 1.3	204640	204673	INS
	pWL046	2067	15345	INS_fragment
	SynII 1.3	**134556**	**134777**	**DEL**
	SynII 1.3	**240777**	**241151**	**DEL**
	SynII 1.3	**260766**	**261157**	**INV**
	SynII 1.3	**295260**	**297781**	**DEL**
	SynII 1.3	540233	540450	DEL
	SynII 1.3	**591520**	**593811**	**INV**
	SynII 1.3	**614390**	**615200**	**DEL**
LWy238	SynII 1.3	212908	212941	INS
	pWL046	15345	2067	INS_fragment
	SynII 1.3	121692	134777	INV
	SynII 1.3	297748	716070	DUP
	SynII 1.3	297748	317917	DUP

	SynII 1.3	690041	720961	DUP
	SynII 1.3	297748	298280	DUP
	SynII 1.3	300615	307392	DUP
	SynII 1.3	307359	317917	DUP
	SynII 1.3	403656	439168	INV
	SynII 1.3	581558	624877	INV
	SynII 1.3	686278	690074	DEL
	SynII 1.3	732447	732697	DEL
LWy258 (LWy238+)	SynII 1.3	212908	212941	INS
	pWL046	15345	2067	INS_fragment
	SynII 1.3	121692	134777	INV
	SynII 1.3	297748	716070	DUP
	SynII 1.3	297748	317917	DUP
	SynII 1.3	690041	720961	DUP
	SynII 1.3	297748	298280	DUP
	SynII 1.3	300615	307392	DUP
	SynII 1.3	307359	317917	DUP
	SynII 1.3	403656	439168	INV
	SynII 1.3	**453559**	**469144**	**INV**
	SynII 1.3	**562415**	**562728**	**INV**
	SynII 1.3	**562415**	**573406**	**Tandem_Dup**
	SynII 1.3	581558	624877	INV
	SynII 1.3	686278	690074	DEL
	SynII 1.3	732447	732697	DEL
LWy259 LWy238-)	SynII 1.3	212908	212941	INS
	pWL046	15345	2067	INS_fragment
	SynII 1.3	121692	134777	INV
	SynII 1.3	**292761**	**298280**	**DEL**
	SynII 1.3	297748	716070	DUP

	SynII 1.3	297748	317917	DUP
	SynII 1.3	690041	720961	DUP
	SynII 1.3	297748	298280	DUP
	SynII 1.3	300615	307392	DUP
	SynII 1.3	307359	317917	DUP
	SynII 1.3	403656	439168	INV
	SynII 1.3	581558	624877	INV
	SynII 1.3	686278	690074	DEL
	SynII 1.3	732447	732697	DEL
LWy239	SynII 1.3	212908	212941	INS
	pWL046	15345	2067	INS_fragment
	SynII 1.3	121692	134777	INV
	SynII 1.3	297748	317917	DUP
	SynII 1.3	297748	298280	DUP
	SynII 1.3	307359	317917	DUP
	SynII 1.3	690041	716070	DUP
	SynII 1.3	403656	439168	INV
	SynII 1.3	581558	624877	INV
	SynII 1.3	686278	690074	DEL
	SynII 1.3	732447	732697	DEL
LWy260 (LWy239+)	SynII 1.3	212908	212941	INS
	pWL046	15345	2067	INS_fragment
	SynII 1.3	121692	134777	INV
	SynII 1.3	297748	317917	DUP
	SynII 1.3	297748	298280	DUP
	SynII 1.3	307359	317917	DUP
	SynII 1.3	690041	716070	DUP
	SynII 1.3	403656	439168	INV
	SynII 1.3	581558	624877	INV

	SynII 1.3	686278	690074	DEL
	SynII 1.3	732447	732697	DEL
	SynII 1.3	**322015**	**395694**	**DUP**
	SynII 1.3	**322015**	**398763**	**INV**
	SynII 1.3	**398730**	**555721**	**DEL**
LWy261 (LWy239-)	SynII 1.3	212908	212941	INS
	pWL046	15345	2067	INS_fragment
	SynII 1.3	121692	134777	INV
	SynII 1.3	297748	317917	DUP
	SynII 1.3	297748	298280	DUP
	SynII 1.3	307359	317917	DUP
	SynII 1.3	690041	716070	DUP
	SynII 1.3	403656	439168	INV
	SynII 1.3	581558	624877	INV
	SynII 1.3	686278	690074	DEL
	SynII 1.3	**581790**	**628076**	**INV**
	SynII 1.3	732447	732697	DEL

* INS: insertion; INV: inversion; DEL: deletion; DUP: duplication; Tandem_DUP: tandem duplication.

* Blue: Source strain/INS; Green: Upregulated SCRaMbLEed strain/ DUP; Red: Downregulated SCRaMbLEed strain/ DEL; Yellow: INV.

* **Bold:** additional genome variations generated by continuous SCRaMbLE compared with source strain.

4.3 Chapter summary, discussion and future work

In this chapter, a pathway SCRaMbLE-in application of the Cre/LoxPsym system was designed and tested for chassis optimization for expressing heterologous pathways in synthetic yeast. With proper induction of CreEBD by β-estradiol and a counter-selection screening strategy, strains with GFP reporter cassettes or either of the two demonstration pathways, β-carotene and violacein, were SCRaMbLEed into the

synthetic chromosome II and selected for further verification. The integration and other genome variations were confirmed by next-generation sequencing and the phenotypes were characterized by chemical analysis. The result proves that SCRaMbLE-in can diversify chemical production by integration to different genomic locations. The violacein pathway SCRaMbLEed-in strains were SCRaMbLEed continuously for further chassis diversification, after which both up- and down-regulated strains were generated with the highest violacein production of 16.8mg/L, almost 17-fold of the HO integrated control strain. A gene complementation experiment was carried out to study the down-regulated strains.

Besides the characterization of SCRaMbLE-in, automated methods for screening strains with higher production after SCRaMbLE have also been tried. However, there are still some difficulties that have not been solved. First, some non-loxP guided integrations to chromosome XV were observed for both GFP cassettes and the violacein pathway, which can add noise and lower the efficiency of FACS screening since it was observed that Chr XV integrated strains have good fluorescence strength or violacein production. Second, not many pathways can be screened by fluorescence based methods like FACS yet. Even for the carotenoids, the fluorescence separation of the strains with different production is not very distinguishable. In addition, the two demonstration pathways are pigment-generating pathways, but for colorless pathways the production can only be characterized by chemical analysis, which is slow, expensive and low-throughput. Therefore, the development and invention of biosensors with reporters for target pathways is also very important for SCRaMbLE-in screening.

The capacity of SCRaMbLE-in tested on SynII is still limited since pathway-relevant genes might be located on other chromosomes. The sensitivity of strains with the violacein pathway and with the β-carotene pathway to SynII SCRaMbLE are different. Therefore, it will be worth trying the SCRaMbLE-in method on different synthetic strains or combined synthetic strains in future for function evaluation. It is probable that the production of violacein or other pathways can be further improved with more genome variation possibilities.

Though the main purpose of SCRaMbLE-in is to screen out strains with higher production, it is worth understanding the genotype and phenotype relationship after SCRaMbLE. The difficulty is that with so many and sometimes complicated genome variations, a huge amount of work will be needed using traditional experimental methods. Therefore, systems biology and computational biology methods should be developed to facilitate the interpretation of genome rearrangements especially when SCRaMbLE or SCRaMbLE-in will be applied to the final synthetic yeast with joint synthetic chromosomes. With the combined high-throughput experimental technologies and computational methods, we can better understand the accelerated evolution of synthetic yeast by SCRaMbLE.

Chapter 5 Gene expression level pathway diversification by recombinase *in vitro*

5.1 Overview and circuit design

In this chapter, the *in vitro* application of recombinases to metabolic engineering at the pathway level will be introduced. Pathway level optimization mainly refers to expression level optimization of catalytic enzymes in the pathway. One popular way of diversifying expression level of genes in a pathway of interest is combinatorial assembly of transcriptional regulation elements like promoters, RBS and terminators with the ORFs of pathway genes (Awan et al., 2017; Y. Guo et al., 2015; Lee et al., 2013; Mitchell et al., 2015). The combinatorial assembly method requires large library construction of DNA circuits. This is used for small pathways with several genes but for large pathways with tens of genes, the coverage of possible combinations can be compromised. Meanwhile, the construction of pathways normally includes several rounds of hierarchical assembly, which can be time- and reagent- consuming especially for large pathways. Here, a recombinase-based integration method was designed for direct integration of transcriptional regulation elements to target genes in a pathway of interest. The advantage of the method is that it can rapidly integrate a library of elements of interest to any DNA fragment with corresponding recombination sites and avoid repeated DNA circuit construction. The recombinase can be simply mixed with DNA and function directly in microcentrifuge tubes. Based on previous work on recombinase *in vitro* function and the recent discovery of new recombinases, Cre, Dre and VCre were chosen to explore the feasibility of this application. The violacein synthetic pathway and β-carotene synthetic pathway were used for demonstrating the functionality of the recombinase toolkit in vitro. A library of characterized yeast promoters was chosen as the transcriptional regulatory elements for integration to regulate key genes in the violacein and β-carotene pathways (Figure 5.1a). VioA, which encodes the first enzyme in the pathway to catalyze the substrate L-Tryptophan, was chosen as the target gene for violacein pathway; CrtI, which encodes the enzyme that transforms

the colorless intermediate phytoene to the yellow neurosporene and red lycopene, was chosen as target gene for the diversification of the β-carotene pathway (Figure 5.1b). Before the final aim of pathway diversification, characterization of the recombinases including purification, quantification of excision and integration rate, and recombination site engineering was performed to evaluate the efficiency of the *in vitro* recombination system. The information of DNA constructs like recombinase expression plasmids, recombination quantification devices, and promoter pathway integration devices are listed in Table 5.1.

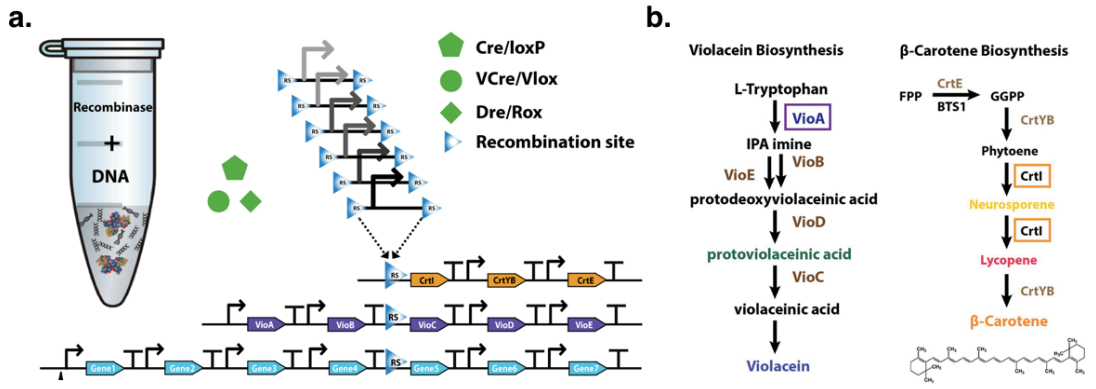


Figure 5.1| Overview of *in vitro* recombinase application. a. Cre/LoxP, VCre/Vlox, and Dre/rox systems were explored for *in vitro* DNA element integration. The recombinases can be mixed directly with DNA and function *in vitro*. The elements of interest in this study are from a YeastFab promoter library with their strength characterized. The integration targets can be upstream of any gene of any pathway of interest. **b.** Violacein and β-carotene synthetic pathways were used for diversification demonstration. VioA, one of the five genes in the violacein pathway, and CrtI, one of the three genes in the β-carotene pathway were chosen as the target genes for promoter integration by recombinases *in vitro*.

Table 5.1 DNA circuit list for recombinase *in vitro* characterization and application

Name	Description	Usage
pWL039	pET28b-Cre	Cre tagging and expression
pWL060	pET28b-Dre	Dre tagging and expression
pWL099	pET28b-VCre	VCre tagging and expression
pWL085	loxPURA3loxP-pRS413	*In vitro* Cre function assay
pWL016	roxURA3rox- pRS415	*In vitro* Dre function assay

pWL022	VloxURA3Vlox-pRS413	*In vitro* VCre function assay
pWL082	pRS415 loxP:RFP:loxP	Excision rate assay
pWL139	pRS415-rox:RFP:rox	Excision rate assay
pWL122	pR415-Vlox-RFP-Vlox	Excision rate assay
pWL083	pRS415 lox71:RFP:lox71	Excision rate assay
pWL084	pRS415 lox66:RFP:lox66	Excision rate assay
pWL086	pRS415 lox15:RFP:lox15	Excision rate assay
pWL087	pRS415 lox510:RFP:lox510	Excision rate assay
pWL088	pRS415 lox17:RFP:lox17	Excision rate assay
pWL089	pRS415 lox71-66:RFP:loxP	Excision rate assay
pWL090	pRS415 lox71-17:RFP:loxP	Excision rate assay
pWL091	pRS415 lox15-66:RFP:loxP	Excision rate assay
pWL109	pRS415 lox15 right-RFP-lox15 right	Excision rate assay
pWL110	pRS415 lox510 right-RFP-lox510 right	Excision rate assay
pWL111	pRS415 lox17 left-RFP-lox17 left	Excision rate assay
pWL112	pRS415 lox15 LR-RFP-loxP	Excision rate assay
pWL113	pRS415 lox510 LR-RFP-loxP	Excision rate assay
pWL114	pRS415 lox17 LR-RFP-loxP	Excision rate assay
pWL115	pRS415 lox m3-RFP-loxP	Excision rate assay
pWL116	pRS415 lox m7-RFP-loxP	Excision rate assay
pWL117	pRS415 lox m3-RFP-loxm7	Excision rate assay
pWL118	pRS415 lox m3-RFP-loxm3	Excision rate assay
pWL119	pRS415 lox m7-RFP-loxm7	Excision rate assay
pWL137	pRS415-lox17m3:RFP:17m3	Excision rate assay
pWL138	pRS415-lox17m7:RFP:17m7	Excision rate assay
pWL202	415 ccdB loxRFPlox	Integration rate assay
pWL203	415 ccdB VloxRFPVlox	Integration rate assay
pWL204	415 ccdB roxRFProx	Integration rate assay
pWL205	415 ccdB 17RFP17	Integration rate assay
pWL206	415 ccdB 17m3RFP17m3	Integration rate assay

pWL207	415 ccdB 17m7RFP17m7	Integration rate assay
pWL208	415 ccdB 17leftRFP17left	Integration rate assay
pWL105	HC_Kan LoxP	Integration rate assay
pWL143	HC_Kan rox single	Integration rate assay
pWL125	HC_Kan Vlox	Integration rate assay
pWL106	HC_Kan lox15	Integration rate assay
pWL107	HC_Kan lox510	Integration rate assay
pWL108	HC_Kan lox71	Integration rate assay
pWL141	HC_Kan 15m3 single	Integration rate assay
pWL142	HC_Kan 15m7 single	Integration rate assay
pWL307	P_{TDH3}-GFP	Expression effect assay
pWL308	P_{TDH3}-loxP-GFP	Expression effect assay
pWL309	P_{TDH3}-lox1517-GFP	Expression effect assay
pWL310	P_{TDH3}-Vlox-GFP	Expression effect assay
pWL311	P_{ACT1}-GFP	Expression effect assay
pWL312	P_{ACT1}-loxP-GFP	Expression effect assay
pWL313	P_{ACT1}-lox1517-GFP	Expression effect assay
pWL314	P_{ACT1}-Vlox-GFP	Expression effect assay
pWL315	$P_{YGR270W}$ GFP	Expression effect assay
pWL316	$P_{YGR270W}$ -loxP-GFP	Expression effect assay
pWL317	$P_{YGR270W}$ -lox1517-GFP	Expression effect assay
pWL318	$P_{YGR270W}$ -Vlox-GFP	Expression effect assay
pWL121	HO-His-YF-RFP acceptor vector	Pathway acceptor vector
pWL160	Single Vlox CrtI β-carotene pathway	Promoter integration pathway
pWL167	pUC19-VloxUraRFPVlox	Promoter loading vector
pWL248	Single Lox15 VioA Violacein pathway	Promoter integration pathway
pWL168	pUC19-lox17UraRFPlox17	Promoter loading vector

5.2 Expression, extraction and purification of recombinases

5.2.1 Circuit design

To express the recombinases in *E. coli*, all recombinase cassettes were cloned into pET-28a (Novagen) to be expressed in-frame with a C-terminal His_6 tag. All recombinase sequences were flanked by NcoI and XhoI restriction sites with start and stop codons removed and a 2bp "gc" was added after the NcoI site for in-frame expression of the recombinase, which encodes an extra glycine after the starting methionine (Figure 5.2). Since glycine is very simple and can fit into both hydrophilic and hydrophobic environment, it should have quite slight effect on the function of a target protein. The source of the recombinases was as follows: Cre was amplified by PCR from a CreEBD expression vector (Lindstrom & Gottschling, 2009); Dre was identified from phage D6, the coding sequence was obtained from DDBJ/EMBL/GenBank database with the accession number AY753669, and Dre was purchased as gBlock from Integrated DNA Technologies (Sauer & McDermott, 2004); VCre was identified from *Vibrio sp. 0908*, the sequence was obtained from GenBank with the accession number ABX77110.1, Vika was identified from *Vibrio coralliilyticus ATCC BAA-450* with the accession number NZ_ACZN01000014.1; SCre was identified from *Shewanella sp. ANA-3* with the accession number ABK50591.1. VCre, Vika and SCre were all purchased from Twist Biosciences. The expression of the recombinases was induced by IPTG in *E. coli* BL21 (DE3).

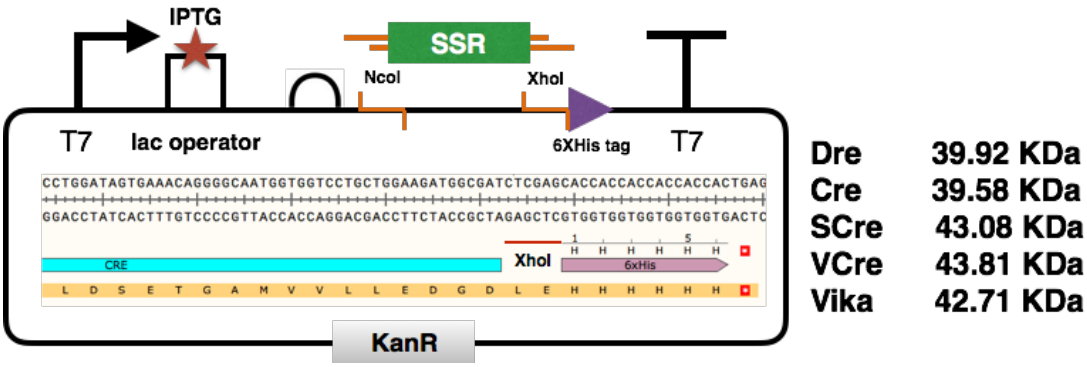


Figure 5.2| Recombinase expression device in *E. coli*. All recombinases were cloned into pET28a for in frame expression with 6X His tag. The expression was induced by IPTG in *E. coli* BL21 (DE3). The sizes of these recombinases are displayed.

5.2.2 Expression and purification

The expression test and purification experiments were done with Laura Tuck from Jon Marles Wright's lab. As proof of principle, only Cre and Dre were tested for expression and purification by affinity chromatography. The sequences of Cre and Dre were initially amplified from the *in vivo* expression circuit pWL078 and pWL079 described in table 3.2. For the expression test, samples were collected before IPTG induction (T0) and after IPTG induction (I-X). After sonication lysis, the cell lysate was centrifuged and both the pellet (I-P) and the supernatant (I-S) were analyzed by SDS-PAGE. The size of Cre and Dre are around 39KDa and when comparing T0, I-P and I-S for proteins of around 39KDa, it was found that all three samples showed abundant proteins at around 40KDa (Figure 5.3a). Though the native proteins make it hard to tell whether the recombinase is expressed, the protein intensity of I-P is stronger than that of T0 for Cre while not much difference is observed for Dre. Not much protein is observed in I-S, which could be the result of incomplete sonication or low solubility. The expression was further confirmed with SDS-PAGE after purification. The result shows that there is indeed a 39KDa protein after purification for Cre while nothing was purified for Dre (Figure 5.3a). One possible reason for failing to express and purify Dre in BL21 (DE3) could be that the sequence of Dre for this round of expression and purification was from the *in vivo* Dre expression device and was yeast codon-optimized.

Therefore, for a second round of expression and purification of the remaining recombinases, all sequences were either from the original phage source strain (for Dre, VCre and Vika) or codon-optimized for *E. coli* for expression by the software GeneDesign (for SCre). Though the sequence of Cre was also not the original one from phage since it had been engineered by error-prone PCR, it was successfully purified and it was not necessary to do any modification.

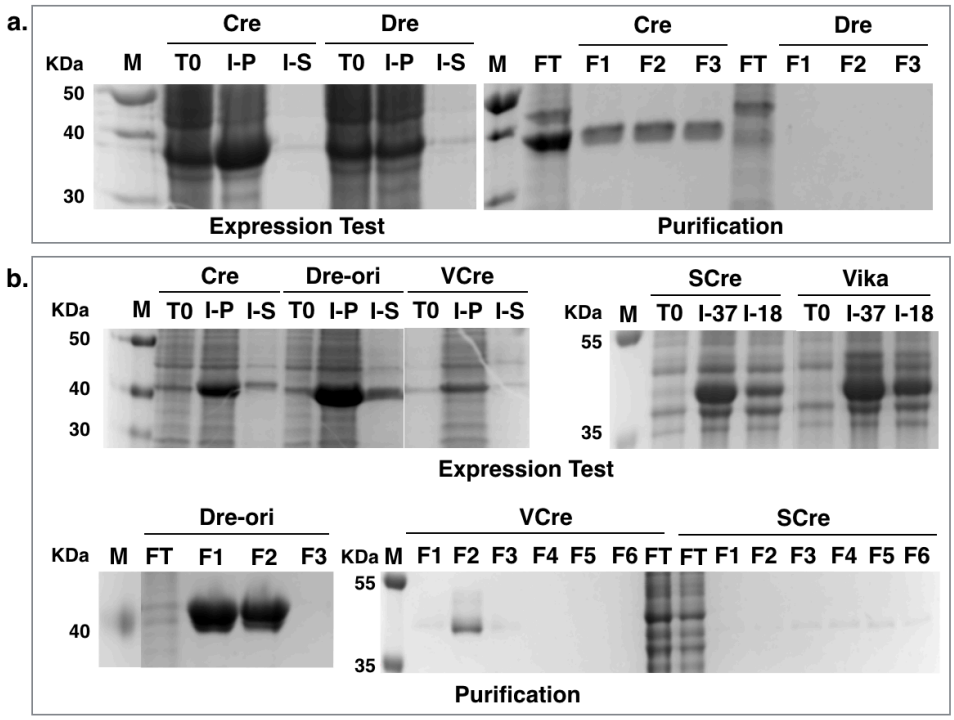


Figure 5.3|Recombinase expression and purification test. a. Primary expression test and purification of Cre and Dre. For Cre, I-P samples have more intense bands at around 39KDa than T0, while for Dre there is not much difference between I-P and T0. The intensity of protein I-S is weak and further modification of the sonication procedure is needed. After protein purification, SDS-PAGE shows that Cre gives a protein band of the right size while there is no band for Dre. **b.** Expression test and purification for Dre-ori, VCre, SCre and Vika. For all five recombinases, I-P shows stronger protein bands at around 40KDa than T0. Cre, Dre-ori and VCre have clear bands in I-S. For SCre and Vika, two induction conditions were compared. I-37 shows a stronger band than I-18 for both recombinases. SDS-PAGE after purification shows that there are protein bands of right size for Dre and VCre, but no band was observed for SCre. M: Marker, protein ladder; T0: Time zero without IPTG induction. I-P: Induced cell pellet; I-S: Induced supernatant; FT: Flow through; F1-6: Collected Fraction 1-6; I-37: Induced under 37°C for 2 h; I-18: Induced under 18°C for 15h. Dre: yeast codon optimized Dre; Dre-ori: original phage sequence of Dre.

From the results of the second round protein expression and purification, we can see that I-P shows more intense protein bands around 39KDa than T0 for all Cre, Dre-ori (original sequence), VCre, SCre and Vika (Figure 5.3b). With condition modification in sonication, the intensity of I-S bands was also increased for Cre, Dre-ori and VCre. For SCre and Vika, two different induction conditions, 37°C for 2 h (I-37) or 18°C for 15h

(I-18), were also compared. It shows that the I-37 samples have relatively higher expression level than I-18 for SCre and Vika. The purification result shows that Dre was successfully purified after switching to the original sequence and VCre was also purified successfully. However, the protein signal during chromatography was very low for SCre and no signal was observed for Vika. The SDS-PAGE result for SCre did not show a band of the correct size and this confirmed the failure of SCre purification (Figure 5.3b). It could be that the two recombinases have poor solubility in water, and solvent optimization could be continued for future purification. However, because the purpose of the study is to select candidates for *in vitro* recombination application, and Cre, Dre and VCre were obtained for function tests, no further investigation on why Vika and SCre were not properly purified was continued.

For final size comparison between Cre, Dre and VCre, SDS-PAGE was done to confirm the purification of the three recombinases (Figure 5.4). The sizes of the Cre and Dre are similar, 39.6KDa and 39.9KDa and VCre is slightly larger, 43.8KDa. It should be noted that the band for Dre is not as clean as the other two, which could indicate that Dre is not very stable and degradation could happen with long-term storage.

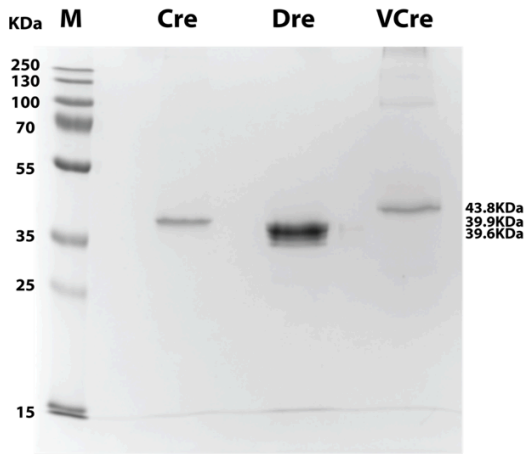


Figure 5.4|SDS-PAGE of purified Cre, Dre and VCre. The size of Cre, Dre and VCre are 39.6KDa, 39.9KDa and 43.8KDa respectively.

5.3 *In vitro* function test and orthogonality test of the recombinases

5.3.1 In vitro functional test strategy

After the recombinases were extracted and purified, the next step was to test whether they had recombination function *in vitro*. Two strategies were used for the function characterization. The first step is to qualitatively evaluate whether recombination between two recombination site happens *in vitro*; the second is to quantitatively calculate the recombination efficiency by applying a red/white color reporting strategy. For the first step verification, the *in vivo* function reporter circuits, the pRS413-site-URA3-site devices, were re-used for *in vitro* function testing. A ScaI digestion map comparison before and after *in vitro* recombination was used to indicate whether recombination happens (Figure 5.5a). This method is very straightforward but it cannot quantitatively describe the recombination efficiency. Therefore, a red/white color reporting strategy was applied to indicate the process of excision or integration between two target sites. Instead of the *URA3* auxotrophic marker, the RFP expression cassette was inserted between two recombination sites (Figure 5.5b). The recombination between two flanking sites will result in the excision of the RFP cassette. When the recombined plasmid is transformed into *E. coli,* the color of the colony reports the recombination event. The excision rate can be calculated by the ratio of white colony number to the total colony number. The excision rate quantification device needs only one plasmid, while the integration rate quantification device needs two plasmids with different antibiotic resistance markers. The first plasmid is similar to the excision rate measuring device, having an *AmpR* cassette and an RFP cassette with two flanking recombination sites but with an additional toxic *ccdB* cassette. The other plasmid has *KanR* and has only one recombination site. The RFP can go through two steps of recombination by first being looped out from the *AmpR* plasmid and then integrated to the *KanR* plasmid. After transformation of the recombination reaction into *E. coli*, the cells were selected on Kanamycin plates and those with integrated RFP were red. By counting the number of red and white colonies, the integration rate could be quantified. Since double transformation of the two original plasmids is possible, the toxic *ccdB* cassette is used

to prevent the survival of "*AmpR-RFP+KanR*" strain and therefore eliminate potential false positive from plasmid co-transformation.

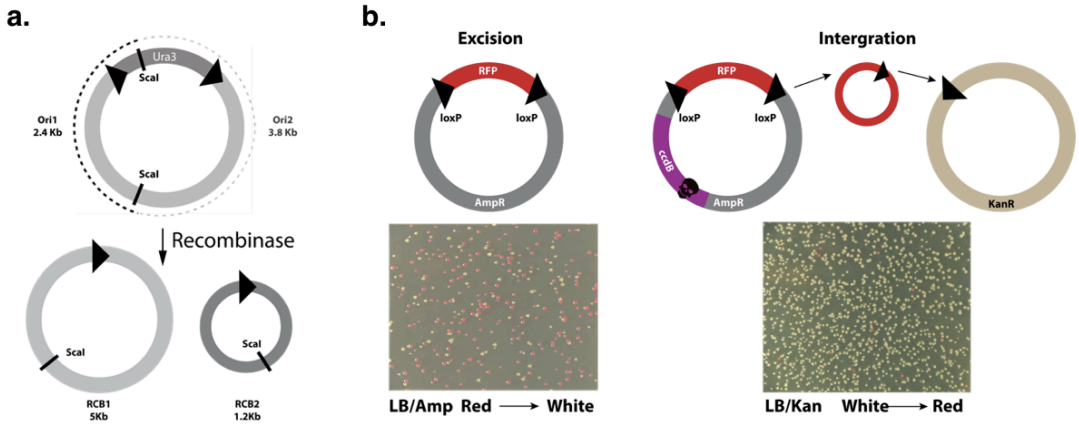


Figure 5.5|*In vitro* recombination function test strategy and circuit design. a. ScaI digestion pattern comparison strategy. The original plasmids should be cut into two bands of 2.4Kb and 3.8Kb. After recombination, the circular DNAs can be cut into two bands of 1.2Kb and 5Kb. **b.** Red/White color reporter devices for recombination efficiency quantification. The excision rate reporting device is composed of a RFP cassette flanked by two recombination sites on an *AmpR* vector. Number of excision events is reported by the number of white colonies. The integration rate reporting device is composed of two plasmids: one is composed of a RFP cassette flanked by two recombination sites and a *ccdB* toxic cassette on an *AmpR* vector; the other is composed of one recombination site on a *KanR* vector. Number of integration event is reported by the number of red colonies on Kanamycin plates.

5.3.2 Orthogonality test between three recombinases in vitro

Three reporter plasmids, pRS413-loxPURA3loxP, pRS413-roxURA3rox and pRS413-VloxPURA3VloxP, were used for the test. The purified Cre, Dre and VCre were mixed with each of the reporter plasmids for the recombination reaction. After 2 h incubation at 37°C, the DNA was purified to remove the recombinase and digested by ScaI. If recombination happens, the digestion pattern of recombined DNA (RCB) will be different from the original template DNA (Ori). The result shows that only Cre with pRS413-loxPURA3loxP, Dre with pRS413-roxURA3rox, and VCre with pRS413-VloxPURA3VloxP have the digestion pattern of RCB bands of 5Kb and 1.2Kb (Figure 5.6). The digestion patterns of the rest are all the same as the negative control, meaning that the functions of the three recombinases are orthogonal. The

recombination efficiency can only be very roughly compared between the three systems and the quantification of the excision and integration rate will be introduced in the next section.

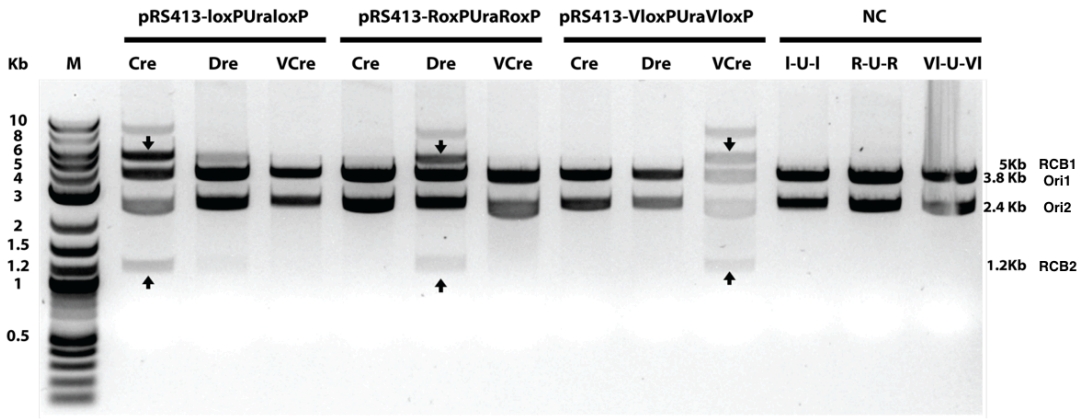


Figure 5.6| Result of digestion assay for *in vitro* recombinase function test. Three reporter plasmids, pRS413-loxPURA3loxP, pRS413-roxURA3rox and pRS413-VloxPURA3VloxP were mixed with Cre, Dre and VCre for recombination reaction. The ScaI digestion shows that only Cre with pRS413-loxPURA3loxP, Dre with pRS413-roxURA3rox, VCre with pRS413-VloxPURA3VloxP have the bands of 5Kb and 1.2Kb recombined DNA. The result proves orthogonal function of most of the recombinases *in vitro* except that Dre has slight leaky function on loxP site. Ori1,2: Original plasmid cutting band; RCB1,2: Recombined DNA cutting band 1,2. NC: negative control, reporter plasmids without recombination. I-U-I: loxPURA3loxP, R-U-R: roxURA3rox, Vl-U-Vl: VloxURA3Vlox.

5.3.3 Excision and integration rate quantification

The recombination efficiency was then tested with the color change strategy. The excision rates show that Cre/loxP has the highest excision rate among the three with 43.5% recombined; VCre/Vlox and Dre/rox system have slightly lower excision rates of 29.7% and 26.2% respectively (Figure 5.7a). The integration rate for Cre/loxP and Dre/rox is around 0.55% and for VCre/Vlox is around 0.3% (Figure 5.7b). The integration rate of natural recombination systems is still at a very low level and further improvement of the integration efficiency is necessary. The quantification of loxP mutant recombination sites for excision and integration rates will be described in the next section.

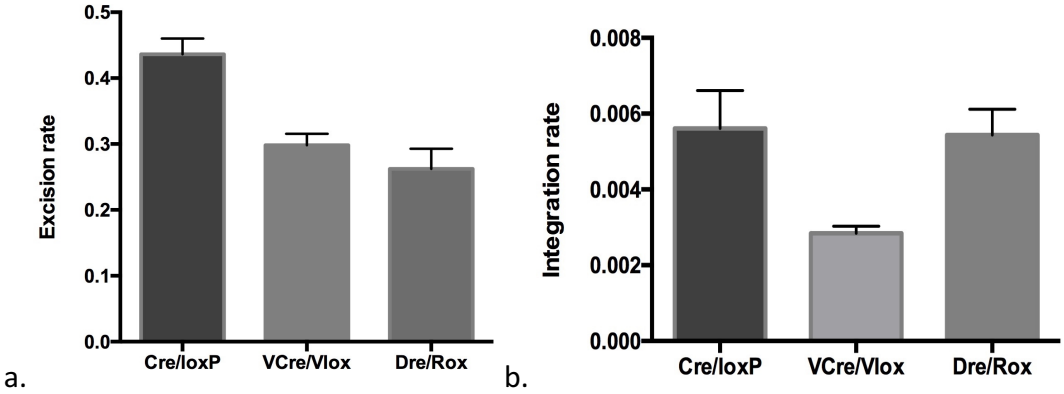


Figure 5.7|Excision rate and integration rate quantification of Cre, VCre and Dre. a. Excision rate quantification. Cre/loxP has the highest excision rate of 43.5%; VCre/Vlox has an excision rate of 29.7%; Dre/rox has an excision rate of 26.2%. **b.** Integration rate quantification. The integration rate of Cre/LoxP is around 0.55%; that of VCre/Vlox is around 0.3%; that of Dre/rox is also around 0.55%. Three *in vitro* reactions were transformed for rate evaluation.

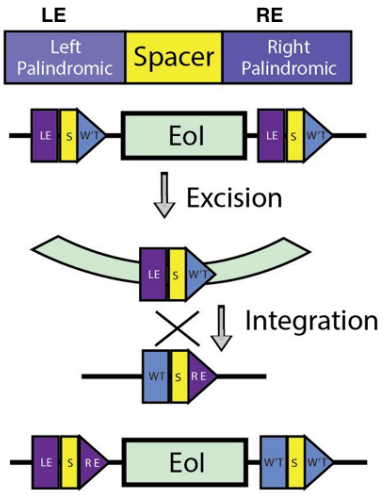
5.4 Engineering of recombination site

5.4.1 Introduction of LE/RE strategy and mutant sites

The recombination process in natural tyrosine recombinases is bi-directional and the integration stability can be compromised by the reverse process. In previous study, the structure of the recombination site was dissected into a spacer region and two palindromic arms and a LE/RE site engineering strategy was developed to improve the integration efficiency *in vivo* (Albert et al., 1995). We also applied the strategy *in vitro* to improve the integration efficiency of "Elements of Interest" (EoI). Two LE mutants were put outside of the EoI and one RE mutant was put in acceptor DNA (Figure 5.8a). The integration between LE and RE sites will result in a LE-RE double mutant site and a wild-type recombination site. The stability of the integration depends on the recombination rate between the LE-RE double mutant and the wild-type site. In this study, the recombination site of the Cre/LoxP system was engineered and quantified to improve the integration rate. In previous studies, five loxP mutant sites, lox71, lox66, loxJT15, loxJT510 and loxJTZ17, were reported to increase the integration rate in ES cells and *E. coli* (K. Araki, Araki, & Yamamura, 1997; J. G. Thomson, E. B. Rucker, 3rd, & J. A. Piedrahita, 2003a). In this study, we compared and

quantified the excision efficiency of these mutants using our quantitative *in vitro* recombination assay. The LE mutants include lox71, loxJT15 and loxJT510 and RE mutants include lox66 and loxJTZ17 (Figure 5.8b). The mutation bases on the palindromic arms of lox71 and lox66 are symmetrical, while the other three are independent mutants. Therefore, we also created the RE mutants of loxJT15 and loxJT510 and the LE of loxJTZ17. The *in vitro* excision assay was carried out both between identical LoxP mutant sites and between the LoxP WT site and LE_RE joint sites. A higher excision rate between identical loxP mutant sites and lower rate between WT and joint mutant site is necessary for better integration. In addition to the palindromic mutants, spacer mutants were also explored for orthogonal integration for future multi-element integration because that spacer region is for homology recognition between different sites. For the proof of principle, M3 and M7 spacer mutant sites were tested (Figure 5.8b).

a.



b.

Palindromic Arm Mutants

Site	Mutation Type	Sequence
LoxP	WT	ATAACTTCGTATAGCATACATTATACGAAGTTAT
lox71	LE	taccgTTCGTATAGCATACATTATACGAAGTTAT
lox66	RE	ATAACTTCGTATAGCATACATTATACGAAcggta
loxJT15	LE	AattaTTCGTATAGCATACATTATACGAAGTTAT
loxJT15 Right	RE	ATAACTTCGTATAGCATACATTATACGAAtaatT
loxJT510	LE	taAcgTTCGTATAGCATACATTATACGAAGTTAT
loxJT510 Right	RE	ATAACTTCGTATAGCATACATTATACGAAcgTta
loxJTZ17 Left	LE	ATAAaTTgcTATAGCATACATTATACGAAGTTAT
loxJTZ17	RE	ATAACTTCGTATAGCATACATTATAgcAAtTTAT

Spacer Region Mutants

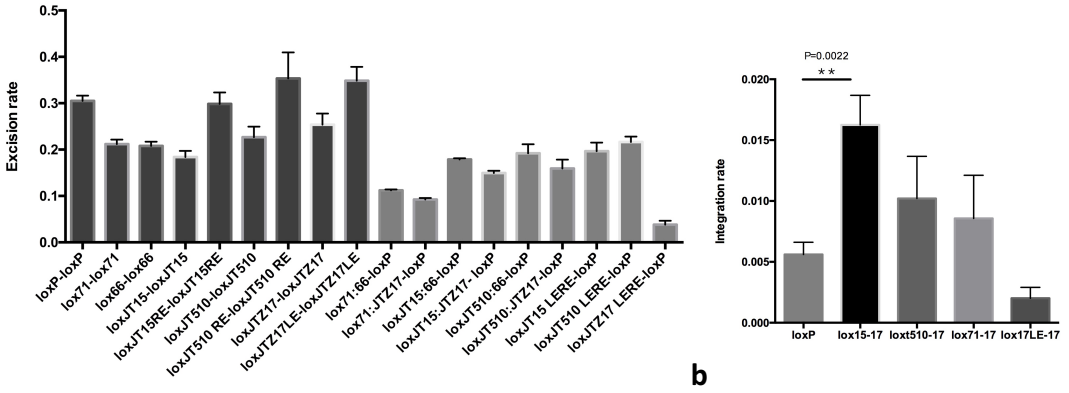
Site	Mutation Type	Sequence
LoxPsym	spacer	ATAACTTCGTATAatgTACATTATACGAAGTTAT
M3	Spacer	ATAACTTCGTATAtggTAttaTATACGAAGTTAT
M7	Spacer	ATAACTTCGTATAttcTAtcTTATACGAAGTTAT

Figure 5.8| Application of LE/RE and spacer mutants of Cre/loxP system *in vitro*. a. Illustration of LE/RE strategy for improving integration efficiency. The EoI is flanked with two LE mutant sites and one RE mutant is put in the acceptor DNA. Through one step of excision and one step of integration, a LE-RE double mutant site and a wild type loxP site are generated. **b.** Sequences of candidate LE, RE and spacer mutant sites of Cre/loxP system that will be tested for function *in vitro*.

5.4.2 Excision and integration rate of LE/RE LoxP mutant

As introduced in 5.1.3, the LE loxP mutant sites (lox71, loxJT15, loxJT510 and loxJTZ17LE) and the RE loxP mutant sites (lox66 or lox71RE, loxJTZ17, loxJT15RE and

loxJT510RE) were quantified for excision rate. The *in vitro* excision assay was carried out both between identical LoxP mutant sites and between the LoxP WT site and LE_RE joint site pairs. A higher excision rate between identical loxP mutant sites and lower rate between WT and joint mutant site is expected to result in more stable integration. The result shows that the excision rates of identical sites were above 20%, generally higher than those of LE_RE and WT pairs, which were below 20% (Figure 5.9a). The excision rate between loxJTZ17LE_17RE and WT Pairs was the lowest and other pairs with low excision rate include lox71_JTZ17 and WT, loxJT510_JTZ17 and WT, loxJT15_JTZ17 and WT, making them good candidates for the integration assay. Interestingly, loxJTZ17 showed a low excision rate in all LE_RE-WT reactions, but gave a high self-excision rate, making it the best candidate site for the EoI loader device. After further tests of integration rate, we found that the integration capability of loxJTZ17 to loxJT15 was the best, around 3-fold that of the loxP WT site (Figure 5.9b). Based on the integration assay of the engineered Cre/LoxP system, we chose the loxJT15 and loxJTZ17 pair as the best candidate for EoI integration.



a **b**

Figure 5.9|Excision and integration rate quantification of LE/RE mutant sites. a. Excision rate comparison between LE/RE mutant sites. The excision rates of LE or RE mutant sites illustrated in dark grey columns are all equal or greater than 20%; the excision rate between LE_RE joint sites and the WT loxP site illustrated in light grey are less than or equal to 20%. The excision rate between loxJTZ17 LE_RE and WT loxP site was the lowest. **b.** Integration rate comparison between candidate LE/RE mutant pairs. The integration rate between loxJT15 and loxJTZ17 is the highest, about 3-fold of that between the WT loxP sites. Error bar represents standard deviation (n=3).

5.4.3 Excision and integration rate of spacer-arm hybrid mutants

Beside the novel use of two orthogonal recombinases *in vitro*, we also studied loxP spacer mutants to generate more orthogonal sites. The spacer region of the recombination site determines the homology recognition during the recombination. Two loxP spacer mutants, m3 and m7, were reported to be orthogonal with loxP WT sites *in vitro*. Our orthogonality test and excision rate calculation of the spacer mutants showed that there was no cross talk between either the m3, m7 and loxP WT sites and the excision rate was similar to that of the loxP WT site (Figure 5.10a).

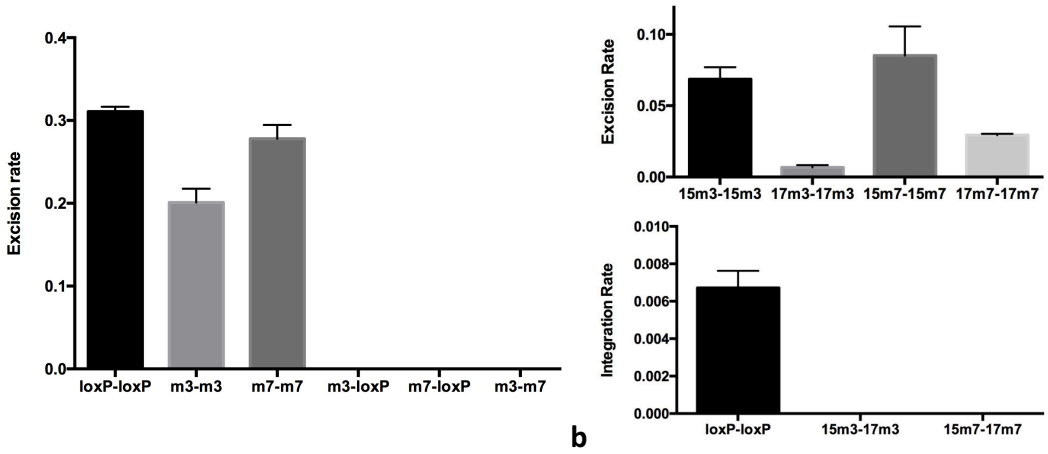


Figure 5.10| Excision and integration rate quantification of spacer mutants and hybrid mutants. **a.** Excision rate of spacer mutants. The excision rate of m7 is similar to the WT loxP site around 30%, and the excision rate of m3 is a bit lower around, 20%. There is no cross talk between loxP, m3 and m7 spacer mutants. **b.** Excision and integration rate of spacer and palindromic hybrid sites. The excision rates of loxJT15m3 and loxJT15m7 are similar, around 7%, while the excision rates of loxJTZ17m3 and loxJTZ17m7 are even lower, around 1% and 3% respectively. The integration rates between loxJT15m3 and loxJTZ17m3, and between loxJT15m7 and loxJTZ17m7 are 0%, which illustrates that the hybrid sites did not work as expected. Error bar represents standard deviation (n=3).

After the excision rate measurement of the spacer mutants, it was decided to combine the spacer mutations and palindromic arm mutations to see whether the integration rate could be improved for these hybrid sites. However, the quantification result shows that the excision rates of the hybrid sites were much lower than loxP WT sites, ranging from 1% to 8% (Figure 5.10b). The integration between loxJT15m3 and loxJTZ17m3 or loxJT15m7 and loxJTZ17m7 hardly happened

(Figure 5.10b), which is disappointing for further application of these sites. It also indicates that our understanding of the recombination sites and function modulation is not yet complete, and the spacer region and palindromic arm cannot be independently mixed and matched for specific function yet.

5.4.4 Design of EoI loader and the YeastFab promoter library

The choice of the EoI is very broad: it can be almost any DNA sequence. To demonstrate the feasibility of its application to metabolic engineering in yeast, a yeast constitutive promoter library was chosen as the EoI library for integration to target pathways. The promoter parts were obtained from the Dai Lab. All parts were standardized to the YeastFab Golden Gate assembly standard, and all were flanked with two BsmBI sites (Y. Guo et al., 2015). A EoI loader was designed for loading these promoters from the library as a first step and will be used for later integration to target genes of interest. In the Design of the EoI loader, a BsmBI-flanked RFP cassette was included for compatible Golden Gate assembly with the promoter library (Figure 5.11a). A *URA3* expression unit was assembled in series with the RFP cassette for integration selection. The *URA3*-RFP cassette was further flanked with the recombination site for the two-step excision-integration reaction. For other standardized libraries, the EoI loader can be adapted correspondingly.

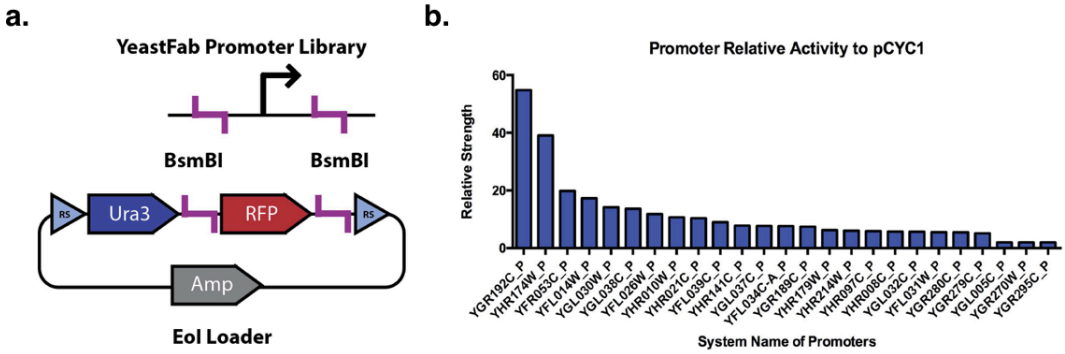


Figure 5.11| EoI loading device and YeastFab promoters for pathway diversification. **a.** Design of EoI loader. The EoI loader is composed of a BsmBI-flanked RFP cassette for the purpose of loading the YeastFab promoters by Golden Gate assembly, a *URA3* cassette for integration selection and two recombination sites for the two-step excision and integration. **b.** The strengths of the candidate promoters have been characterized and standardized to P_{CYC1}. Twenty-five promoters with strength from 2-50 were used for pathway integration.

The function of the promoter library has been partially characterized with fluorescent proteins. The activity of the promoters was standardized against the activity of pCYC1 in the YeastFab paper, with activity ranging from 0 to 50 (Y. Guo et al., 2015). From these characterized promoters, 25 were chosen with three different ranges of strength, including “>10”, “2-10”, and “<2” to test their ability to diversify gene expression and further diversify chemical production (Figure 5.11b). Detailed information about the promoters is listed in Table 5.2.

Table 5.2 Standardized YeastFab promoters

	System name	Promoter activity relative to P_{CYC1}
1	YGR192C_P	54.89
2	YHR174W_P	39.14
3	YFR053C_P	19.88
4	YFL014W_P	17.34
5	YGL030W_P	14.21
6	YGL038C_P	13.73
7	YFL026W_P	11.84
8	YHR010W_P	10.73
9	YHR021C_P	10.45
10	YFL039C_P	9.03
11	YHR141C_P	7.81
12	YGL037C_P	7.73
13	YFL034C-A_P	7.67
14	YGR189C_P	7.49
15	YHR179W_P	6.26
16	YHR214W_P	6.08
17	YHR097C_P	5.95
18	YHR008C_P	5.77
19	YGL032C_P	5.72
20	YFL031W_P	5.56

21	YGR280C_P	5.54
22	YGR279C_P	5.15
23	YGL005C_P	2.03
24	YGR270W_P	2.01
25	YGR295C_P	2.01

5.5 Application of the recombinases for pathway diversification *in vitro*

5.5.1 Design and construction of EoI acceptor for pathway diversification

To achieve EoI integration, there must be a target recombination site on the acceptor DNA. For the purpose of diversifying gene expression using different promoters, the recombination site should be located upstream of the gene to be regulated. This can be achieved by adding a recombination site tail during PCR amplification of the target gene. To make the toolkit more flexible for broader use, the recombination sites were also made as standardized bioparts in the YeastFab standard. As shown in Figure 5.12, the RFP cassette was surrounded by either two loxP, two rox or two VloxP sites respectively. These recombination site flanking RFP cassettes were put into the YeastFab HCKan_P vector as the promoter libraries for Golden Gate construction of the pathway. Taking β-carotene pathway construction as an example, three rounds of Golden Gate assembly were included. In step 1, all site-RFP cassettes and promoters were assembled to HCKan_P vector, the coding genes were assembled to HCKan_O vector and the terminators were assembled to HCKan_T vector by BsaI Golden Gate assembly. In step 2, the site-RFP cassette/promoter, CDS and terminators were assembled into POT vectors by BsmBI Golden Gate assembly. In step 3, for those having the recombination site, a round of *in vitro* recombination was done to get rid of the RFP with a single site left upstream of the gene. In the final step, the POT cassettes were assembled to the EoI acceptor vector by BsaI Golden Gate assembly. The EoI acceptor vector has two HO locus homology arms and a *HIS3* selection marker for genome integration.

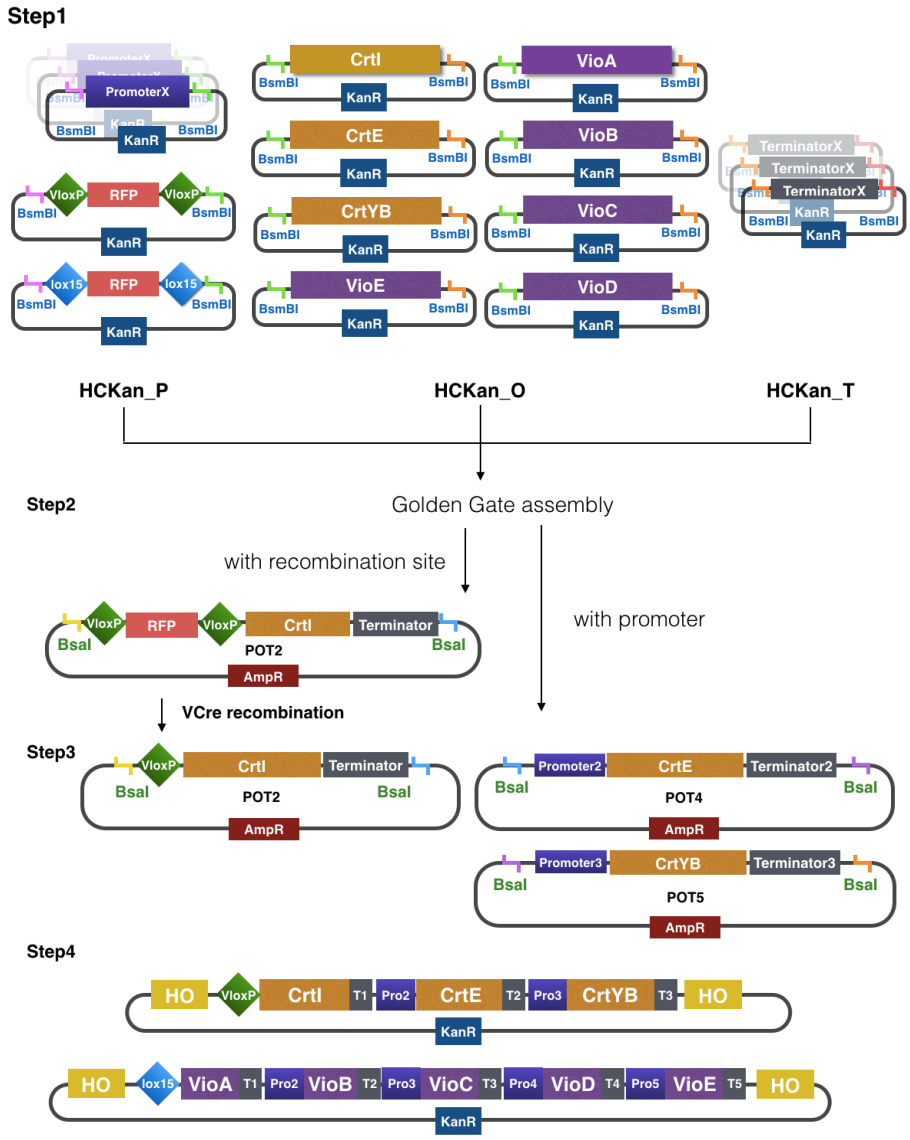


Figure 5.12| Construction strategy of recombination site containing pathway circuit. The construction of pathway circuit includes three rounds of Golden Gate assembly. Step1: recombination sites flanking RFP and constitutive promoters were put into HCKan_P receptor vector; CrtI, CrtE, CrtYB, VioA-E coding sequences were put into HCKan_O receptor vector; terminators were put into HCKan_T vector. Step2: RFP/promoter element, CDS and terminator were assembled into POT unit in different vectors by BsmBI Golden Gate assembly. Step3: the unit with recombination site flanking RFP went through one round of in vitro excision by recombinase; unit with constitutive promoter will skip to step4. Step4: POT expression units and recombination site containing units were inserted into final receptor vector by BsaI Golden Gate assembly. All BsmBI sites and BsaI sites were YeastFab standard. The final assembly for Violacein pathway with recombination site was shown in step 4 with step 2 and step 3 omitted.

5.5.2 Pathway diversification workflow using the *in vitro* recombinase toolkit

In summary, the standardized procedures for pathway diversification with the recombinase toolkit include loading of YeastFab promoter library, two-step excision-integration *in vitro* recombination reaction between loaded EoI and EoI acceptor and a final transformation to host yeast. Strains are then selected on SC-His-Ura plates and picked for phenotype and genotype analysis (Figure 5.13). The phenotype assays include fitness tests and product quantification; the genotype includes determination of which promoters had integrated.

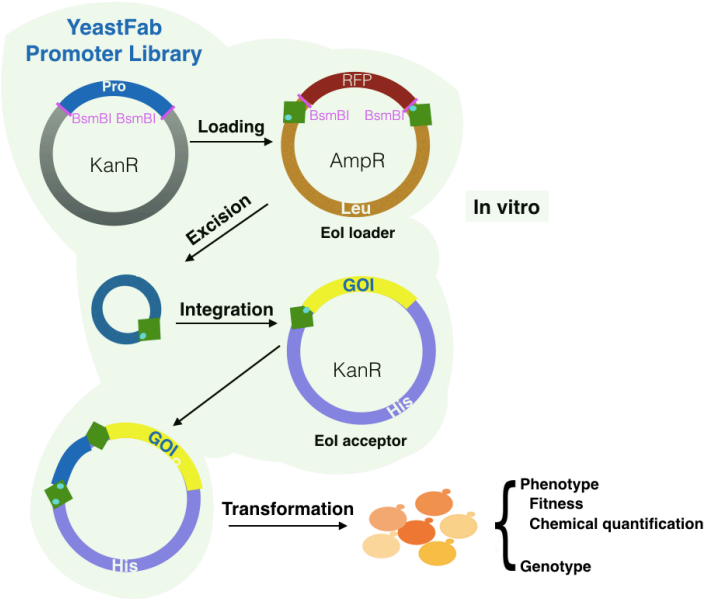


Figure 5.13| Pathway diversification workflow by *in vitro* promoter integration. The *in vitro* application workflow of recombinases on pathway diversification includes promoter loading to EoI loader, in vitro incubation for promoter excision and integration into EoI acceptor, and transformation to yeast for pathway production, follows further phenotype and genotype characterization. EoI: element of interest.

5.5.3 Effect of recombination site on gene expression

Before finally applying the system to promoter integration into gene of interest, I also did an assay to evaluate the potential effect of the recombination site on gene expression. Considering the palindromic structure of the recombination site, a hairpin could form with the opening of double-strand DNA, and affect transcription efficiency. In the YeastFab library, the right BsmBI overhang of the EoI loader was “GATG”, therefore the translation of the protein will start from the LoxP or Vlox site.

In order to put the site in frame with the GoI, the 2bp sequence"TG" was put between loxP or Vlox and the start codon of the target gene.

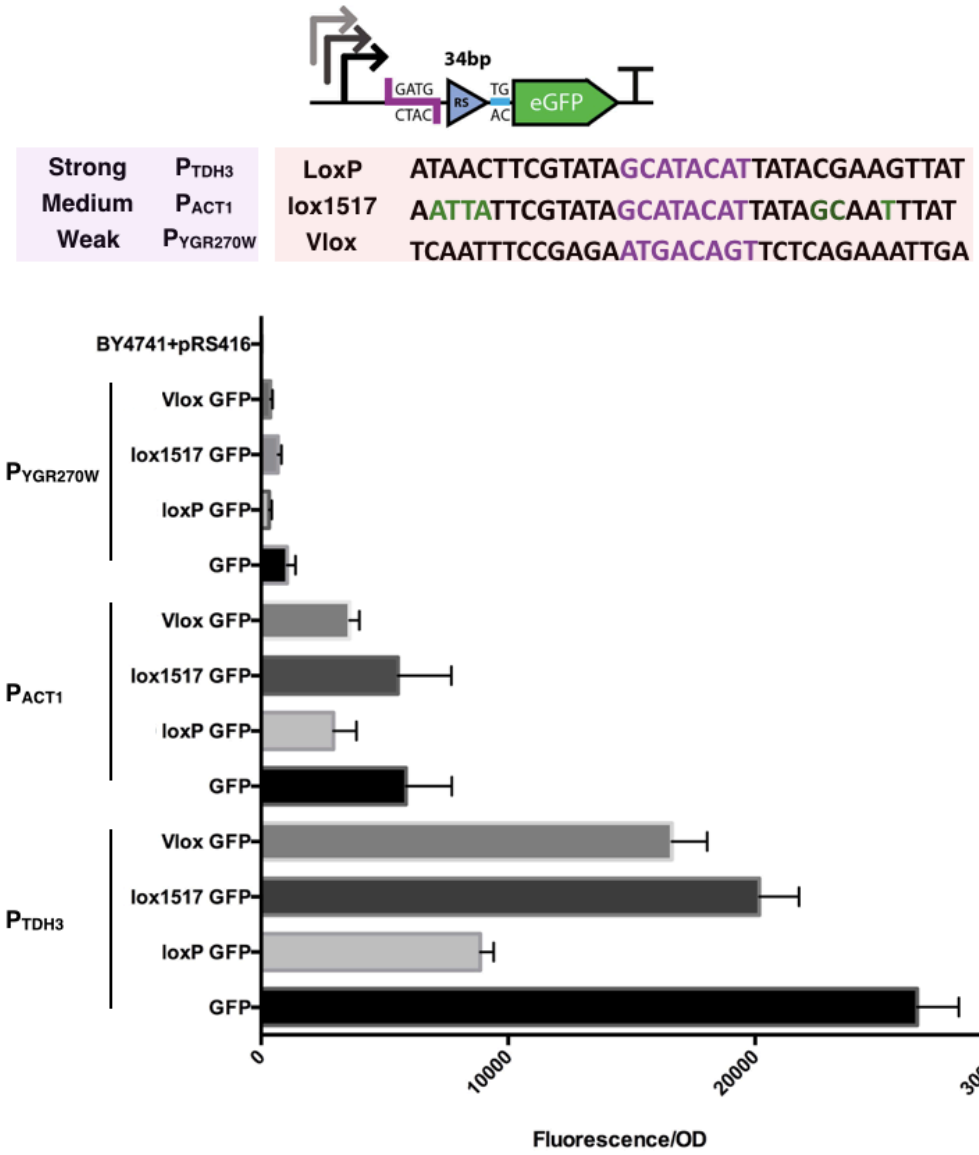


Figure 5.14| Effect of recombination sites on gene expression. LoxP, VloxP, and loxJT15-JTZ17 were inserted between promoters and the eGFP gene. Since the overhang of BsmBI contains the ATG start codon and the site is 34bp, the 2bp sequence "TG" was inserted after the recombination site for in-frame expression of GFP. Strong promoter P_{TDH3}, medium promoter P_{ACT1} and a weak promoter $P_{YGR270W}$ were used to compare the similarity of recombination site effects under different promoters. It was found that recombination sites indeed have a negative effect on GFP expression and the effects are similar with different promoters. Double mutant site loxJT15-JTZ17 has the smallest effect on gene expression; VloxP has medium effect; loxP has the strongest negative effect.

We tested the effect of the recombination site by putting the selected three recombination sites, loxP, loxJT15-JTZ17 and Vlox, between the promoter and the GFP open reading frame. Three promoters with different transcription strengths were used to see whether promoter strength can affect the result. We found that the recombination sites have a similar effect pattern on gene expression under no matter strong, medium or weak promoters (Figure 5.14). Among the three-recombination sites, loxP has an obvious negative effect on gene expression; Vlox has lighter negative effect than the loxP site; the double mutant site loxJT15-JTZ17 has least effect on gene expression. This could be explained by the asymmetric structure of the double mutant site. Therefore, the LE/RE strategy can not only improve the integration efficiency but also minimize the transcriptional effect of the recombination site on gene expression for the promoter integration application.

5.5.4 Proof of principle experiment by promoter integration to *URA3* gene

As a proof of principle promoter integration experiment, promoters P_{TDH3} and P_{URA3} were used for integration to the *URA3* gene. Since there is an "ATG" sequence in the loxP site, there will be a 7-amino acid tag added to the N terminus of the Ura3 protein (Figure 5.15a). The promoters P_{TDH3} and P_{URA3} were assembled to the EoI loader as described in the previous section. After 7h *in vitro* recombination, the reaction was directly transformed into yeast and screened on SC-His-Ura plates (Figure 5.15b). After plate selection, the plasmids were recovered from clones for sequencing confirmation. It was found that there are both P_{TDH3} and P_{URA3} integrated isolates (Figure 5.15c).

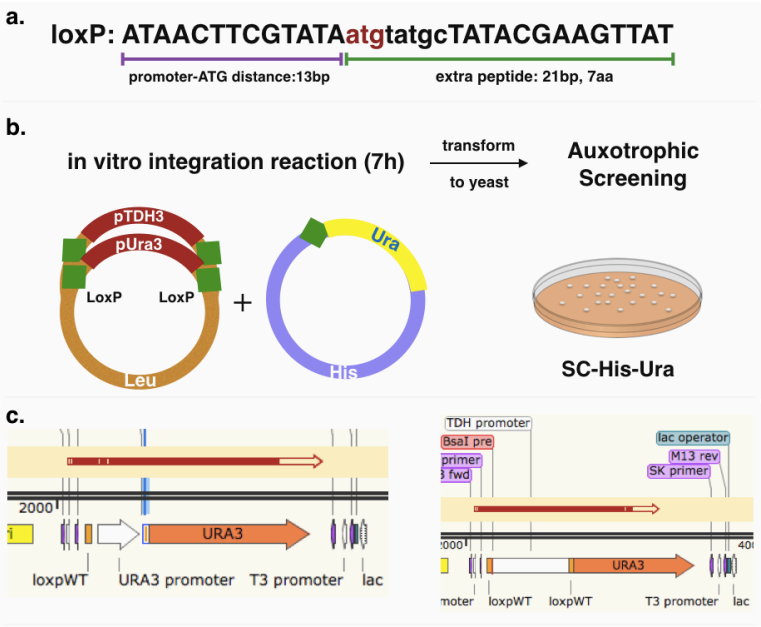


Figure 5.15|Promoter integration test by *URA3* gene activation. a. Impact of loxP sites on the protein sequence of *URA3* gene. The existence of "atg" and the assembly method will result in a 7 amino acid peptide tag being added to the N terminus of Ura3. **b.** Integration screening workflow. After 7h incubation with the recombinase, the reaction was transformed into yeast and selected on SC-His-Ura plates. **c.** Sequencing verification of P_{TDH3} and P_{URA3} integration.

5.5.5 Pathway diversification and characterization

Based on the characterization results, Cre/loxJT15-JTZ17 and VCre/Vlox were selected for pathway diversification with the YeastFab promoter library. Two benchmark pathways, β-carotene and violacein, were used for demonstration. After promoter loading, *in vitro* recombination and yeast transformation, strains with different pathway circuits were selected on SC-His-Ura. As we can see for the β-carotene pathway, this *in vitro* promoter integration generated yeast colonies with a range of different colors and colony sizes, indicating diversified yields or chemical composition of the carotenoid production (Figure 5.16a). The integration and genotype of the promoter were confirmed by sequencing (Table 5.3). The fitness and chemical analysis by HPLC or LCMS of the promoter-integrated variants were further compared (Figure 5.16b-d).

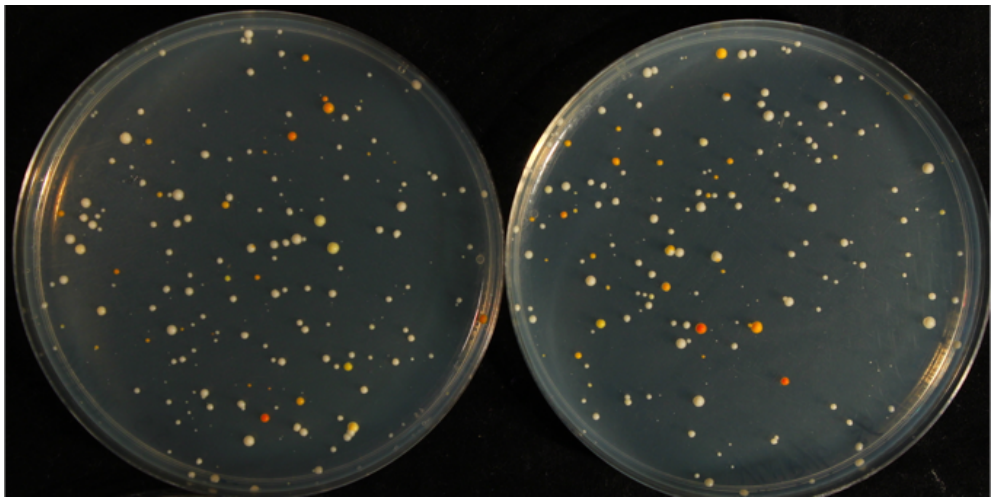
Table 5.3 List of strains with verified promoter integration

Strain ID	Strain	Description	Genotype
LWy068	By4741	P_{YPT1} Promoter integration VloxCrtI	leu2Δ0 met15Δ0 LYS2
LWy069	By4741	P_{RPL22B} Promoter integration VloxCrtI	leu2Δ0 met15Δ0 LYS2
LWy070	By4741	P_{STE2} Promoter integration VloxCrtI	leu2Δ0 met15Δ0 LYS2
LWy071	By4741	P_{HXK1} Promoter integration VloxCrtI	leu2Δ0 met15Δ0 LYS2
LWy072	By4741	P_{ACT1} Promoter integration VloxCrtI	leu2Δ0 met15Δ0 LYS2
LWy073	By4741	P_{HSP12} Promoter integration VloxCrtI	leu2Δ0 met15Δ0 LYS2
LWy076	By4741	P_{TDH3} Promoter integration VloxCrtI	leu2Δ0 met15Δ0 LYS2
LWy087	By4741	P_{SCW4} Promoter integration VloxVioA	leu2Δ0 met15Δ0 LYS2
LWy091	By4741	P_{RPL27} Promoter integration lox1517VioA	leu2Δ0 met15Δ0 LYS2
LWy092	By4741	P_{AGA2} Promoter integration lox1517VioA	leu2Δ0 met15Δ0 LYS2
LWy093	By4741	P_{HXK1} Promoter integration lox1517VioA	leu2Δ0 met15Δ0 LYS2
LWy094	BY4741	P_{PNC1} Promoter integration lox1517VioA	leu2Δ0 met15Δ0 LYS2

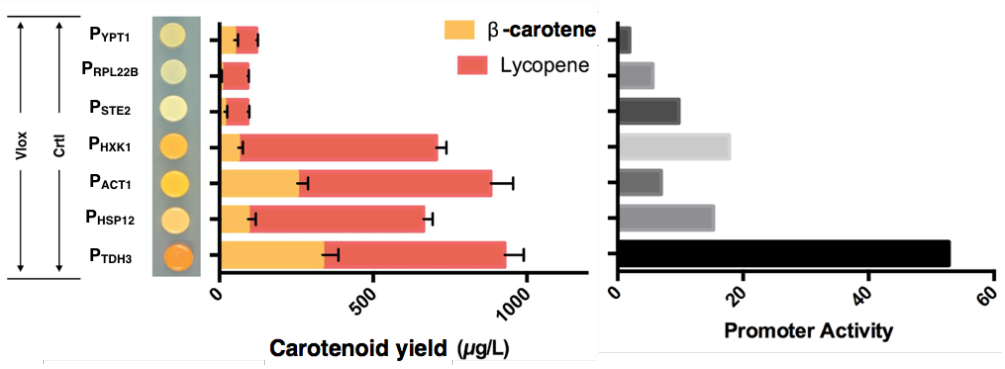
For the β-carotene pathway, the production of lycopene and β-carotene were quantified by LCMS. The results and activity of the integrated promoters are listed in Figure 5.16b. The result shows that *CrtI* integrated with the strongest promoter P_{TDH3} has the highest carotenoid production. For the other variants, the carotenoid production is not in proportion with the activity of the promoters; for example the P_{ACT1}-integrated strain has similar yield with P_{TDH3}-integrated strain, while the P_{RPL22B}-integrated strain has the lowest β-carotene yield even though P_{RPL22B} is not the weakest promoter. In addition, since the function of *CrtI* was reported to catalyze transformation of phytoene to lycopene, theoretically the ratio of β-carotene to Lycopene should be relatively stable(Figure 5.1b). However, the result shows that different ratios of β-carotene to lycopene were generated. It is possible that there are still unknown functions of *CrtI* or there are unknown effects of the intermediates on the promoters. Besides production difference, the fitness of the integrated strains was also different (Figure 5.16c). It was found that the low production strains with P_{YPT1} and P_{RPL22B} have longer doubling time and lower saturation cell density; the high production strains with P_{TDH3}, P_{HSP12} and P_{ACT1} have shorter doubling time and higher

saturation point; the carotenoid production of strains with P_{STE2} and with P_{HXK1} were quite different, while the fitness for the two strains are similar. Such fitness difference could be the result of metabolic burden from the accumulated intermediates in β-carotene pathway and the flux could be affected by promoter choices, while more work has to be done to understand the mechanisms behind it.

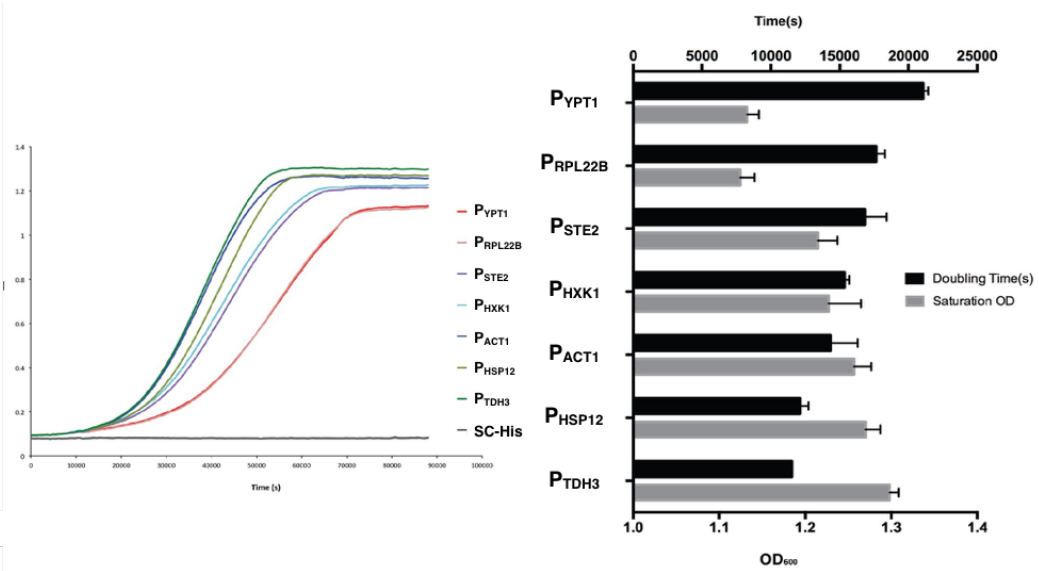
a

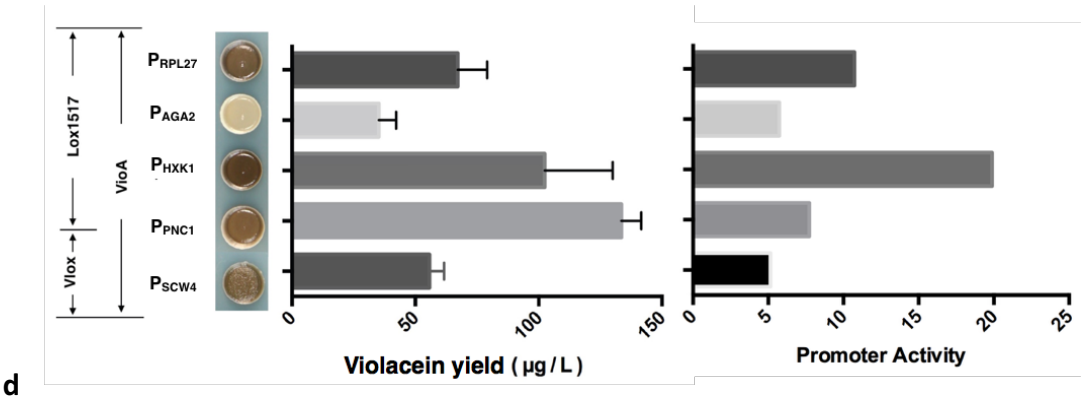


b

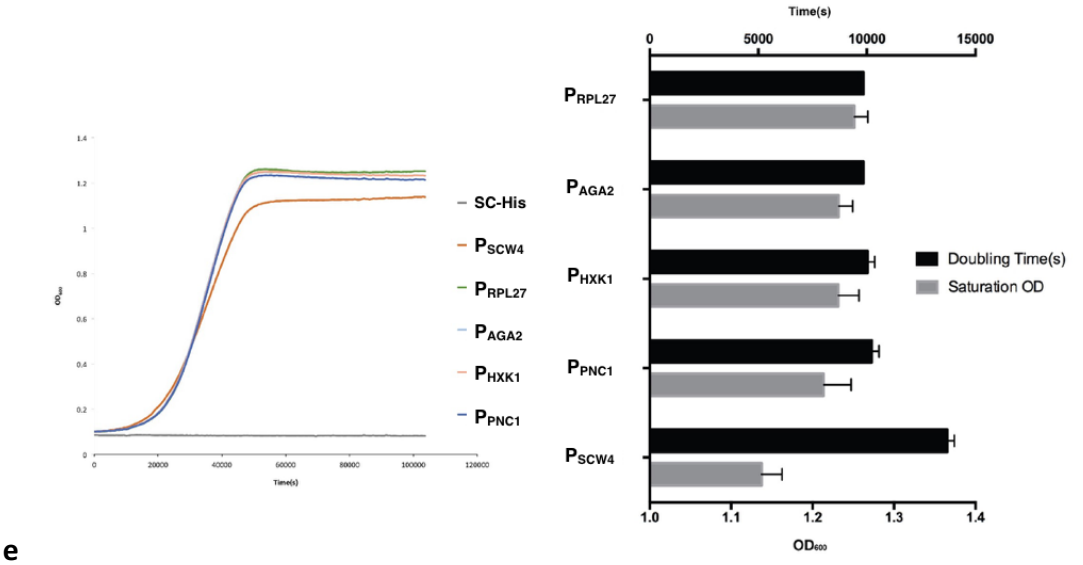


c





d



e

Figure 5.16|Pathway diversification and phenotype-genotype analysis. a. Transformation results of the *in vitro* integration reactions of promoters into the β-carotene acceptor pathway. **b.** Carotenoids production and promoter activity comparison between the strains with *CrtI* gene targeted β-carotene pathway. **c.** Fitness comparison between the strains with *CrtI* gene targeted β-carotene pathway. **d.** Violacein production and promoter activity comparison between strains with *VioA* gene targeted violacein pathway. **e.** Fitness comparison between the promoter integrated strains with violacein pathway.

The relationship of promoter activity with the chemical production is different for different pathways. When comparing violacein production with the activity of the integrated promoters, the degree of correlation is higher than that of β-carotene pathway with only one exception, the strain with medium promoter P_{PNC1} has a higher production (Figure 5.16d). The fitness of the promoter integrated strains of

the violacein pathway is more similar to each other with only one slower strain integrated with P_{SCW4} (Figure 5.16e).

5.6 Chapter summary, discussion and future work

In this chapter, the *in vitro* application of recombinases for metabolic engineering is described. Three out of five recombinase candidates, Cre, Dre and VCre were successfully extracted and purified. The excision rate and integration rate of the three natural systems were quantified by the red/white color ratio strategy. To improve the integration efficiency, the LE/RE mutated sites of the Cre/loxP system were tested *in vitro*. The integration rate of the loxJT15 and loxJTZ17 pair was proven to be 3-fold higher than the natural loxP site. A YeastFab promoter library was chosen as integration element for pathway diversification. With the *in vitro* Cre/loxJT15-JTZ17 and VCre/Vlox systems, the β-carotene and violacein pathways were successfully diversified. The results prove that the idea of element integration by recombinase activity *in vitro* and its application to metabolic engineering is feasible. There are some limits of the methods. Since the integration rate is still fairly low even after LE/RE engineering, a selection marker for the integration of the promoters is necessary to get rid of the background. This is feasible for single gene diversification, but the need for systematic multiple regulation of genes in a pathway is increasing and combinatorial assembly is only the current solution. The *in vitro* integration rate has to be further improved to achieve multiple integration and more orthogonal sites or systems should be discovered or designed for this purpose. There are two aspects that can be optimized or attempted. One is to further investigate the working conditions of the recombination system *in vitro*, for example improving on the protein storage buffer and recombination reaction buffer. The other is to replace the tyrosine recombinase with a serine recombinase, also known as integrase, for the purpose of integration (Olorunniji et al., 2017; Stark, 2017). However, the length of the recombination site is even longer and the effects on gene expression have to be carefully evaluated as well for the purpose of promoter integration.

Chapter 6 Synthetic Natural Product platform for screening new antibiotics

6.1 Overview of the SynNP platform for discovery of new antibiotics

In previous chapters, short pathways such as the β-carotene and violacein synthesis pathways were used for demonstrating the function of the recombinase tools. However, these two pathways have already been well studied and it is worth working on designing and engineering more pathways for real-life applications. To extend my PhD research on the application of synthetic biology and metabolic engineering for heterologous expression of brand new pathways in *S. cerevisiae*, I joined the SynNP three-partner consortium to produce and screen for new antimicrobials active against tuberculosis.

6.1.1 Workflow of pathway diversification screening

Antibiotics can be biologically derived chemicals from microbes and plants or be chemically designed and synthesized. Naturally synthesized chemicals are the main source of antibiotics. Amongst natural sources, *Streptomyces* in the phylum of *Actinobacteria* produces over two-thirds of the clinically useful antibiotics. Therefore, the species in *Streptomyces* are good starting resources for isolating new antibiotic synthetic pathways. Through genome mining and literature search and review, many biosynthetic gene clusters (BGCs) from *Streptomyces*, *Nocardia* and many other microbe genera were selected as candidates for expression and engineering in yeast (Figure 6.1a). Five pilot natural product classes were focused on for proof of principle construction and production tests, including non-ribosomal peptide synthases based pathways (NRPSs), ribosomally synthesized and post-translationally modified peptides based pathways (RiPPs), polyketide synthetases based pathways (PKSs), nucleosides and flavonoids. In each of the classes, an exemplar compound which has been proven to be active against *Mycobacterium tuberculosis* (viomycin, nocathiacin I, kirromycin, capuramycin and naringenin) was first tested with the construction and diversification methods and the screening assays. All the biosynthetic genes (BSGs)

were codon optimized for *S. cerevisiae* using programming scripts of Almer from the Tyers group.

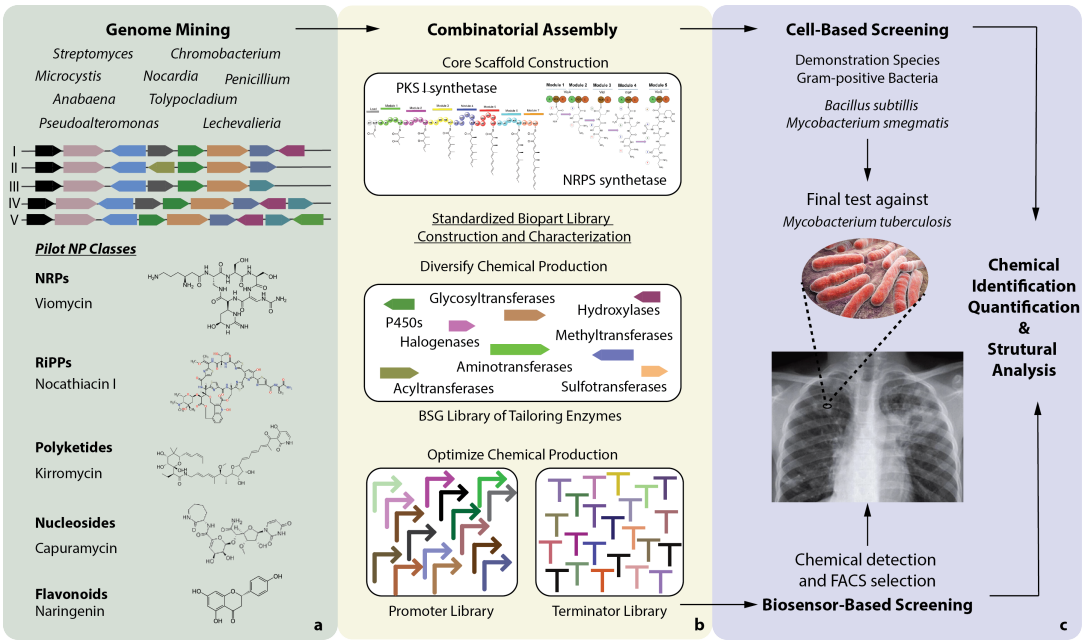


Figure 6.1| Workflow of SynNP pathway construction, diversification and characterization. a. Genetic library collection for SynNP pathways by genome mining and literature search. **b.** Diversification and optimization of SynNP pathways by combinatorial assembly of tailoring enzymes or regulatory elements. **c.** Screening and characterization assays for antibiotic production and activity from the SynNP pathway library. NP: natural product. NRPs: non-ribosomal peptide synthases-based pathway. RiPPs: ribosomally synthesized and post-translationally modified peptides based pathway.

Since many of these pathways are quite large, with either many genes or with very large genes, different assembly strategies and hierarchical assembly methods are needed for pathway construction. For pathway diversification, two combinatorial assembly strategies were designed: combinatorial assembly of tailoring enzymes and combinatorial assembly of regulatory elements. The precursor molecules first need to be processed by multi-domain synthetases to synthesize the basic chemical scaffolds in the NRPs or PKS pathways, or transformed by posttranslational modification (PTM) to form a central core ring as in the RiPPs pathways (Figure 6.1b). Once the synthesis of the basic core chemical scaffold is confirmed, the pathway can be processed for chemical diversification by combinatorial assembly with different

tailoring enzymes such as group-transferases, halogenases, P450, etc. The initial heterologous expression level of each gene might not be favorable for the metabolic flux balance, therefore expression level optimization is very likely needed for later optimization of antibiotic production. The YeastFab promoter and terminator libraries were again applied for combinatorial assembly for the SynNP project.

Since libraries of synthetic pathways will be constructed for antibiotic selection, high-throughput functional assays are necessary for preliminary screening of pathways for antibiotic activity. Both traditional cell-based screening and artificial biosensor-based screening can be used for detecting product formation (Figure 6.1c). For cell-based assay, ideally *Mycobacterium tuberculosis* should be the test strain. However, because *M. tuberculosis* is of biosafety level-BSL3 and only one partner from the consortium has this capability, not all assays can be done with *M. tuberculosis*. Substitute BSL-1 species like *Bacillus subtilis* and *Mycobacterium smegmatis* were therefore used for initial antibiotic function testing since they are all gram-positive bacteria. For the biosensor-based assay, the sensor can be either specific compound-responsive reporters like riboswitches in yeast or *in vitro*, or be stress sensing reporters in the pathogens. When some candidates are screened out, further chemical extraction, quantification and structural analysis of the metabolites can be run for confirmation.

Considering that the scale of the project is large, different missions were allocated to different groups and different people oversaw different pathways. My contribution to the project included the design and characterization of a set of DNA assembly vectors compatible with YeastFab standards and yeast *in vivo* assembly standards, construction and characterization of the RiPPs nocathiacin I pathway, and setting up a violacein producing yeast as positive control for the cell-based assay.

6.1.2 Strategy for screening new antibiotics

For the cell-based assays of antibiotics screening, a pathogen overlay assay and a disk diffusion assay were applied (Figure 6.2). In the overlay assay, yeast cells with a SynNP pathway are cultured on plates until large colonies are formed; then a

pathogen-medium mixture is poured on top of the yeast culture; if the yeast colony is secreting antimicrobials above the minimal inhibition concentration (MIC) for the pathogen, a zone of inhibition around the yeast colony will be formed (Figure 6.2a). For some large chemicals or chemicals without proper transporter proteins in yeast membranes, the capability of the overlay assay can be quite limited. In this case, the disc diffusion assay with yeast extract is a good alternative method for antibiotic activity test (Figure 6.2b).

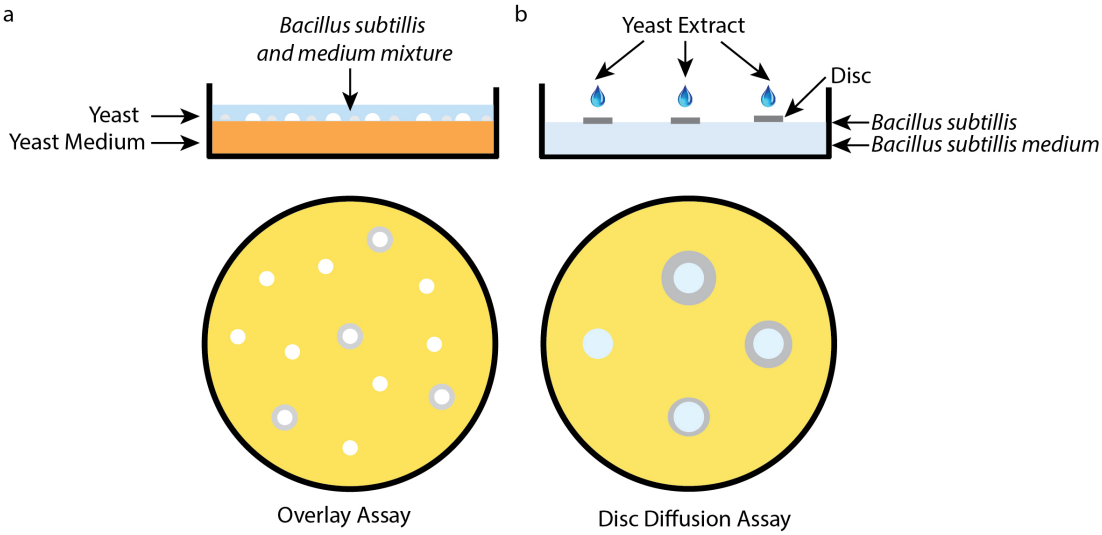


Figure 6.2|Screening assays for antibiotic selection. a. Illustration of the cell-based overlay assay. **b.** Illustration of the disc diffusion assay.

6.1.3 General workflow for pathway construction and the design of tool vectors

The construction methods for a SynNP pathway include 3 rounds of alternating BsaI and BsmBI Golden Gate assembly and one yeast *in vivo* assembly. In the Golden Gate assembly, the type IIs restriction enzymes BsaI and BsmBI were used and these sites were removed from the BSGs if present to facilitate the assembly strategy. In step 1, the promoters, BSGs and terminators were PCR-amplified and respectively assembled to HCKan_P, HCKan_O and HCKan_T vectors using BsaI; in step 2, transcription units (TUs) were constructed by assembling promoter, BSG and terminator in a specific POT vector with BsmBI Golden Gate assembly; in step 3, up to six TUs are assembled to a specific pYFASS vector as multi-TU clusters by BsaI golden gate assembly; in the final step, TU clusters on different pYFASS vectors were

joined together to a YAC vector by yeast homologous recombination-based assembly (Figure 6.3).

The HCKan vectors and POT vectors are published in Dai and colleagues' YeastFab work (Y. Guo et al., 2015). The YAC vector is from our collaborator Tyers group. I designed and constructed the pYFASS vectors to enable compatibility between the Golden Gate and yeast assembly standards. In the design of the pYFASS vectors, a *ccdB* counter selection marker was flanked with two BsaI sites that are compatible with YeastFab assembly; the cassette was surrounded by two 100bp arms for yeast assembly; outside the cassette lay two BsmBI sites, for linearization of the cassette from the backbone. With different choices of POT vector, the pYFASS vectors are able to accommodate one to six TUs to form multi-TU gene clusters (Figure 6.3 step3). The number of the BSGs in the pathway determines how the POT vectors will be used and the large nocathiacin I pathway, which has 26 genes, will be introduced as an example. Since in this SynNP project nocathiacin I is the pathway with most BSGs, five pYFASS vectors were constructed to accommodate up to 30 BSGs. Six orthogonal yeast assembly arms were used for assembling 5 multi-TU clusters (Figure 6.3 step 4). The BSG capacity can be theoretically infinitely extended by adding more homology arms for yeast assembly. After the linearization of the YAC vector by BamHI + XbaI + SacI and pYFASS-multi-TUs by BsmBI, all DNA fragments were co-transformed into yeast and selected on SC-His-Leu plates with optional 5-FOA counter-selection.

In each assembly step, there are quality control (QC) procedures to make sure the assembly is correct and no mutation is introduced. The assembly QC procedures in each step are summarized in Table 6.1. Besides QC of DNA assembly, the expression and function characterization of each BSG can be carried out in step2 since the POT vectors are centromeric plasmids. The combinatorial assembly of regulatory elements can be carried out at step two. One more advantage of the multi-TU cluster assembly is that the BSGs can be categorized into groups for modular study. For instance, BSGs involved in core scaffold synthesis can be grouped to the same cluster, while tailoring enzymes can be grouped to other clusters and combinatorial assembly of the tailoring enzymes can be separately achieved.

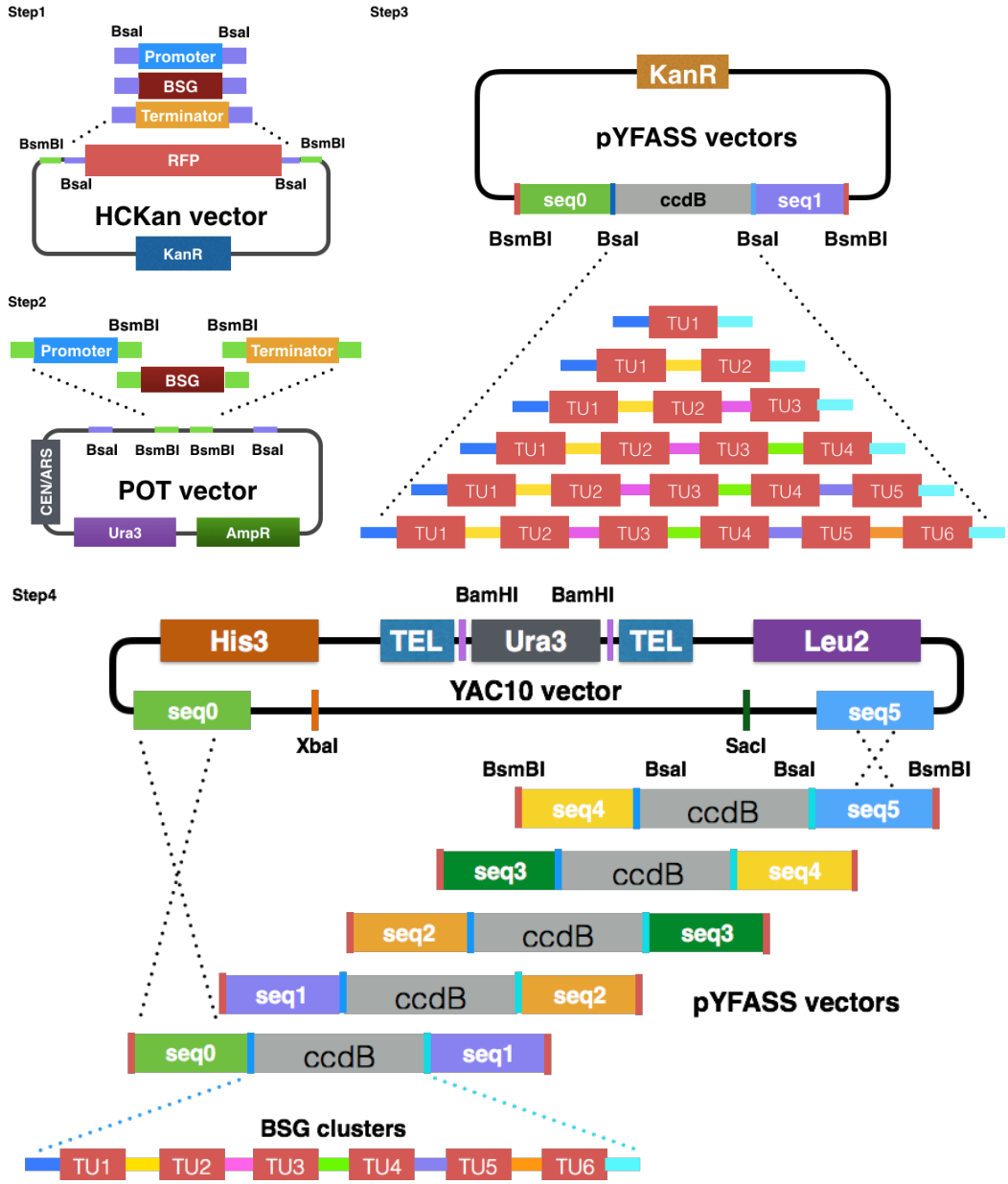


Figure 6.3 | Workflow of pathway construction. Step 1: Promoters, BSGs, terminators are standardized with YeastFab BsaI overhangs and assembled to HCKan_P, HCKan_O and HCKan_T vectors respectively. **Step 2:** Promoter, BSG and terminator are assembled with a specific POT vector to form a transcription unit (TU) by BsmBI Golden Gate assembly. BSG expression and function can be characterized at this step. **Step 3:** One to six TUs can be assembled to the pYFASS vector to form a multi-TU cluster by BsaI Golden Gate assembly. Each pYFASS vector contains two 100-bp arms for yeast *in vivo* assembly. **Step 4:** For pathways with up to 30 BSGs, five pYFASS vectors can be used to accommodate multi-TU clusters and the clusters are finally assembled to a YAC vector by *in vivo* yeast assembly. The YAC vector should be linearized by BamHI + XbaI + SacI and the clusters should be linearized by BsmBI before co-transformation to yeast.

Table 6.1 Pathway construction strategy and QC methods

Step	Part, Assembly and Vector	Enzyme	Screen and Quality Control
1	PCR amplify BSGs Golden Gate Assembly HCKan_O vector	BsaI	Colony PCR and Sequencing verification
2	HCKan_O assembled BSGs Single TU Golden Gate Assembly POT vector	BsmBI	Colony PCR and digestion
3	POT assembled TUs Multi-TU Golden Gate Assembly pYFASS vector	BsaI	Digestion
4	Multi-TU clusters Yeast YAC Assembly YAC10 vector	BsmBI BamHI XbaI, SacI	Junction colony PCR, BSG colony PCR and sequencing

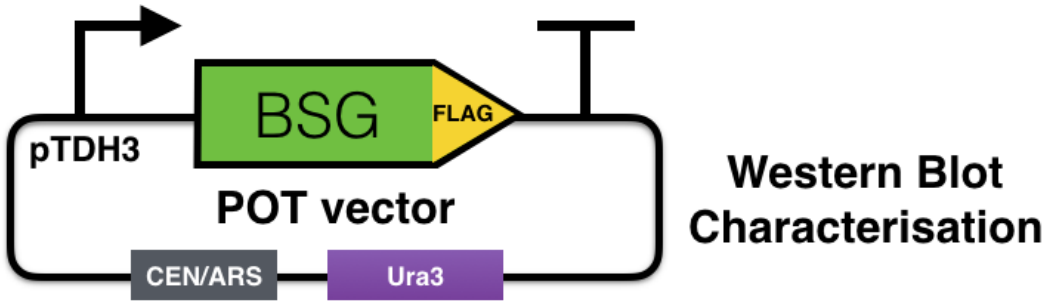
6.1.4 RiPPs exemplar pathway--nocathiacin I and construction strategy

The glycosylated thiopeptide nocathiacin I was identified from the *Nocardia* sp. ATCC202099 and reported to have antibiotic activity against a variety of gram-positive bacteria including *M.tuberculosis* (Ding et al., 2010; Pucci et al., 2004). The nocathiacin I pathway is an exemplar pathway of the RiPPs and is composed of 26 genes. The functions of the genes have been predicted by aligning with a homologous RiPPs pathway, the nosiheptide (NOS) pathway(Ding et al., 2010). The NocM gene, like the NosM gene, encodes the precursor peptide for the pathway. Since the expression level of NocM determines the quantity of substrate for the pathway, it is better to get it highly expressed to achieve a higher production of the final nocathiacin I. Therefore, the strongest promoter, P_{TDH3}, was used to drive the transcription of NocM, and four copies of the NocM transcriptional unit were assembled. Thus, 29 genes were assembled in total, with 6 genes assembled to each of pYFASS vectors 1-4 and the remaining 5 genes to pYFASS5 (Figure 6.4a). To avoid

unwanted recombination between NocM transcription units, the four copies of the NocM were allocated to different pYFASS vectors. Six genes that are related to sugar decoration were assembled together to the pYFASS2 to facilitate future engineering of these tailoring enzymes. Except for the NocM TU, the rest of the TUs all have different promoter and terminator pairs to avoid any unspecific recombination (Figure 6.4a). At the same time, all BSGs were tagged with FLAG for yeast expression assay by western blot. These FLAG-tagged genes were separately assembled with the P_{TDH3} promoter for strong expression in the POT2 vector to enhance the western blot signal (Figure 6.4b).

a.

Position	BSG	Length	POT vector	YFpASS vector	Promoter	Terminator
1	NocP	952	POT2	YFPASS1	YGR248W_P	YAL056C-A
2	NocB	1261	POT4	YFPASS1	YHR008C_P	YAL011W
3	NocM1	202	POT6	YFPASS1	YGR192C_P	YAL042C-A
4	NocQ	676	POT8	YFPASS1	ACT1 5'UTR	YAL056W
5	NocA	508	POT10	YFPASS1	YHR087W_P	CIT1 3'UTR
6	NocT	1165	POT11	YFPASS1	YGL037C_P	FUM1 3'UTR
7	NocS1	1231	POT2	YFPASS2	YHR214W_P	YAL063C
8	NocS2	766	POT4	YFPASS2	YGL032C_P	YAL017W
9	NocS3	1297	POT6	YFPASS2	PGK1 5'UTR	YAL047W-A
10	NocS4	1039	POT8	YFPASS2	RPS2 5'UTR	YAL063C-A
11	NocS5	1483	POT10	YFPASS2	YHR021C_P	YAL010C
12	NocS6	1189	POT11	YFPASS2	YFL034C-A_P	ENO2 3'UTR
13	NocU	1291	POT2	YFPASS3	YHR097C_P	YAL067W-A
14	NocR	577	POT4	YFPASS3	YFL031W_P	YAL024C
15	NocM2	202	POT6	YFPASS3	YGR192C_P	YAL042C-A
16	NocV	1195	POT8	YFPASS3	ZEO1 5'UTR	YAL068C
17	NocC	1333	POT10	YFPASS3	YFL039C_P	YAL016W
18	NocD	1045	POT11	YFPASS3	YGR260W_P	YAL041W
19	NocF	1492	POT2	YFPASS4	YFR053C_P	YAR008W
20	NocG	1948	POT4	YFPASS4	YHR174W_P	YAL031C
21	NocM3	202	POT6	YFPASS4	YGR192C_P	YAL042C-A
22	NocH	1834	POT8	YFPASS4	YHR021C_P	YAR014C
23	NocI	1285	POT10	YFPASS4	YHR128W_P	YAL023C
24	NocK	940	POT11	YFPASS4	YGR189C_P	YAL047C
25	NocE	2611	POT2	YFPASS5	YFL014W_P	YAL002W
26	NocL	1189	POT4	YFPASS5	YGR280C_P	YAL036C
27	NocM4	202	POT6	YFPASS5	YGR192C_P	YAL042C-A
28	NocN	1315	POT8	YFPASS5	YGR279C_P	YAL003W
29	NocO	994	POT9	YFPASS5	YFL026W_P	ACS1 3'UTR



b.

Figure 6.4|Assembly and expression characterization strategy for the RiPPs exemplar pathway nocathiacin I. a. Assembly strategy for nocathiacin I. The total assembly for nocathiacin includes 4 copies of NocM plus the remaining 25 genes, so 29 TUs in total. Except the four NocM TU, all driven by P_{TDH3}, the other TUs were all composed of different promoter and terminator pairs from the YeastFab library. Each of pYFASS1-4 accommodates 6 BSG TUs and pYFASS5 contains 5 BSG TUs. The NocM TU were allocated to different pYFASS vectors and all sugar modification genes NocS1-6 were all assembled to pYFASS2. **b.** Expression characterization circuit for BSGs. All BSGs were tagged with FLAG for western blot expression test. All the tagged BSGs were assembled with the same strong promoter, P_{TDH3}, to achieve a strong western blot signal. The TUs were assembled to POT2 vector and sent for test at step2.

6.1.5 List of DNA circuits

The information of pYFASS vectors, assembled nocathiacin gene clusters and FLAG tagged BSGs are shown in Table 6.2. The yeast strains with nocathiacin I pathway are shown in Table 6.3.

Table 6.2 Key plasmids for pathway construction and characterization

Name	Description	Usage
pWL181	pYFASS1 (overhang YF compatible ACCT-TGAG)	Multi-TU assembly
pWL182	pYFASS2 (overhang YF compatible ACCT-TGAG)	Multi-TU assembly
pWL183	pYFASS3 (overhang YF compatible ACCT-TGAG)	Multi-TU assembly
pWL184	pYFASS4 (overhang YF compatible ACCT-TGAG)	Multi-TU assembly
pWL185	pYFASS5 (overhang YF compatible ACCT-TGAG)	Multi-TU assembly
pWL325	NocP NocB NocM1 NocQ NocA NocT TUs on pYFASS1	Noc multi-TU cluster1
pWL326	NocS1 NocS2 NocS3 NocS4 NocS5 NocS6 TUs on pYFASS2	Noc multi-TU cluster2

pWL327	NocU NocR NocM2 NocV NocC NocD TUs on YFPASS3	Noc multi-TU cluster3
pWL290	NocF NocG NocM3 NocH NocI NocK TUs on YFPASS4	Noc multi-TU cluster4
pWL328	NocE NocL NocM4 NocN NocO TUs on YFPASS5	Noc multi-TU cluster5
pWL384	P_{TDH3}-nocP FLAG tag	
pWL385	P_{TDH3}-nocS1 FLAG tag	
pWL386	P_{TDH3}-nocB FLAG tag	
pWL387	P_{TDH3}-nocQ FLAG tag	
pWL388	P_{TDH3}-nocA FLAG tag	
pWL389	P_{TDH3}-nocS2 FLAG tag	
pWL390	P_{TDH3}-nocS3 FLAG tag	
pWL391	P_{TDH3}-nocS4 FLAG tag	
pWL392	P_{TDH3}-nocS5 FLAG tag	
pWL393	P_{TDH3}-nocS6 FLAG tag	
pWL394	P_{TDH3}-nocT FLAG tag	
pWL395	P_{TDH3}-nocU FLAG tag	Expression test by western blot
pWL396	P_{TDH3}-nocR FLAG tag	
pWL397	P_{TDH3}-nocV FLAG tag	
pWL398	P_{TDH3}-nocC FLAG tag	
pWL399	P_{TDH3}-nocD FLAG tag	
pWL400	P_{TDH3}-nocE FLAG tag	
pWL401	P_{TDH3}-nocF FLAG tag	
pWL402	P_{TDH3}-nocG FLAG tag	
pWL403	P_{TDH3}-nocH FLAG tag	
pWL404	P_{TDH3}-nocI FLAG tag	
pWL405	P_{TDH3}-nocK FLAG tag	
pWL406	P_{TDH3}-nocL FLAG tag	
pWL407	P_{TDH3}-nocN FLAG tag	
pWL408	P_{TDH3}-nocO FLAG tag	

Table 6.3 Yeast strains with nocathiacin I pathway

Strain ID	Strain	Description	Genotype
LWy098	By4741	YAC10-Nocathiacin I version 1 (point mutation in NocN)	leu2Δ0 met15Δ0 LYS2 ura3Δ0 his3Δ1
LWy218	BY4741	YAC10-Nocathiacin I version 2	leu2Δ0 met15Δ0 LYS2 ura3Δ0 his3Δ1

6.2 Results

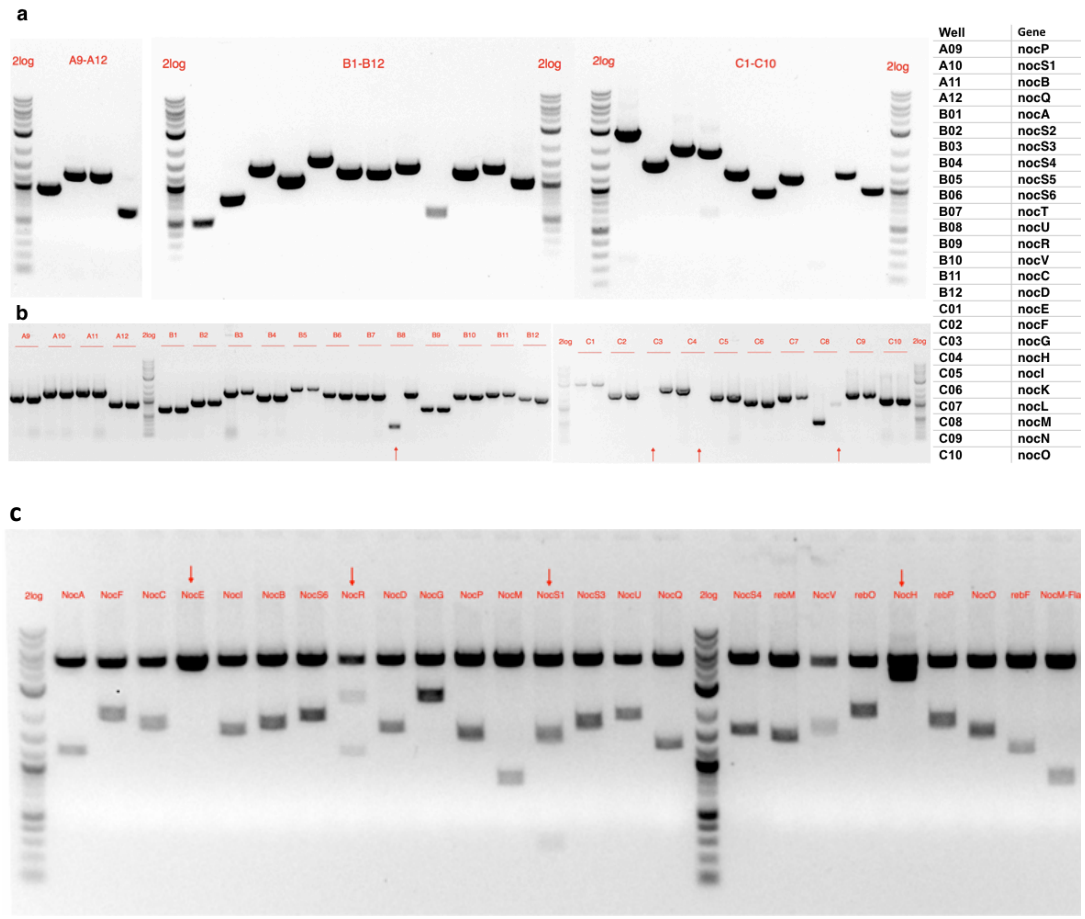
6.2.1 Construction of pYFASS assembly vectors

The pYFASS vectors were based on the HCKan_T vectors and the construction includes one round of Golden Gate assembly and one round of Gibson assembly. In the first step, the 100bp yeast assembly arms were standardized with BsaI overhangs and assembled into HCKan_T by Golden Gate assembly with EcoIV restriction site between the two arms. In the second step, the *ccdB* cassette was amplified and tailed with the ACCT-TGAG YeastFab-compatible BsaI overhangs by PCR, and was inserted to the EcoIV site by Gibson assembly. The reaction was transformed into ccdB Survival™ 2 strain for *ccdB* gene cloning. The *ccdB* counter-selection marker, the YeastFab-compatible BsaI sites and the homologous arms for yeast *in vivo* assembly were put together as designed, and since the HCKan_T vector naturally contains a pair of BsmBI sites flanking the arms, the construction of the pYFASS vectors was complete.

6.2.2 Construction of the nocathiacin I pathway

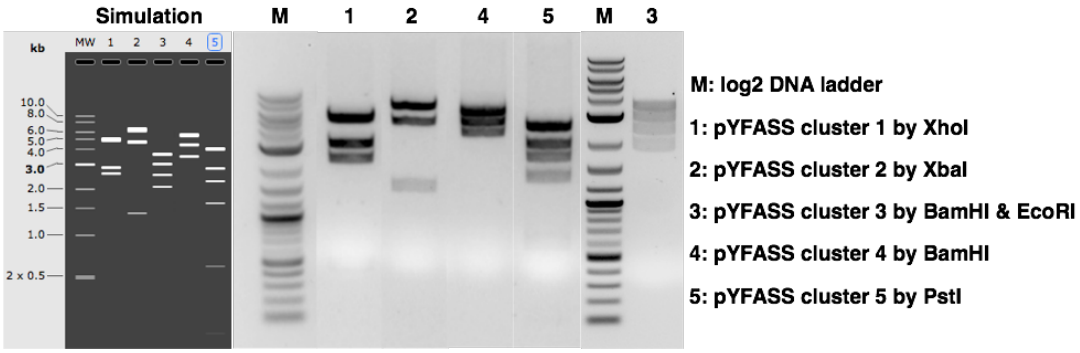
PCR amplification of all BSGs was successful (Figure 6.5a). After assembly with HCKan_O vector, two colonies from each of the genes were picked for colony PCR screening. The result shows that at least one of each has the right size (Figure 6.5b). The HCKan_BSGs were sent for sequencing confirmation before POT assembly to avoid introduction of any mutation in the PCR process. Once the HCKan_BSGs passed the sequencing test, they were assembled with the proposed promoter and terminator pairs. A similar colony PCR was used for clone screening and BsaI digestion of the POT-TU plasmids was used to double-check the correctness of the assembly

(Figure 6.5c). At the beginning, there were some negative results which did not show the correct digestion pattern. After trouble-shooting, it was found that this was because not all BsaI and BsmBI sites were removed from the promoters and terminators to avoid affecting the function of the elements, but they reduced the efficiency of the Golden Gate assembly. Then those with either BsaI or BsmBI sites were substituted with other promoters from the YeastFab library to avoid such problems. When the correctness of each POT-TU assembly was confirmed, they went forward to multi-TU cluster assembly with pYFASS vectors. The correctness was checked by digestion using appropriate restriction enzymes (Figure 6.5d). Then the multi-TU clusters were released from the vector by BsmBI and the YAC10 vector was linearized by BamHI, XbaI and SacI (Figure 6.5e). All multi-TU cluster bands and the two YAC10 bands (4kb and 6.5kb) were gel extracted and recovered. All fragments were co-transformed to BY4741 for yeast *in vivo* assembly.



Well	Gene
A09	nocP
A10	nocS1
A11	nocB
A12	nocQ
B01	nocA
B02	nocS2
B03	nocS3
B04	nocS4
B05	nocS5
B06	nocS6
B07	nocT
B08	nocU
B09	nocR
B10	nocV
B11	nocC
B12	nocD
C01	nocE
C02	nocF
C03	nocG
C04	nocH
C05	nocI
C06	nocK
C07	nocL
C08	nocM
C09	nocN
C10	nocO

d



e

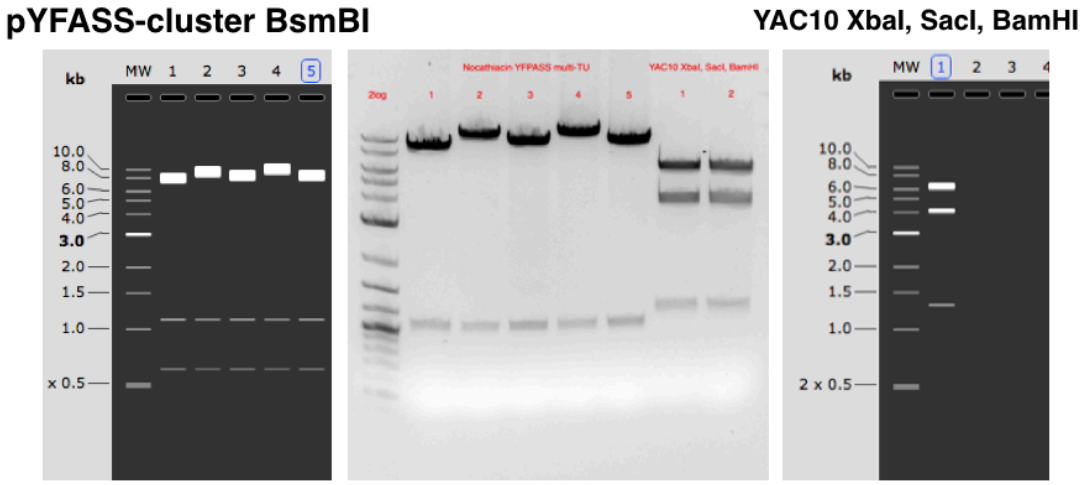
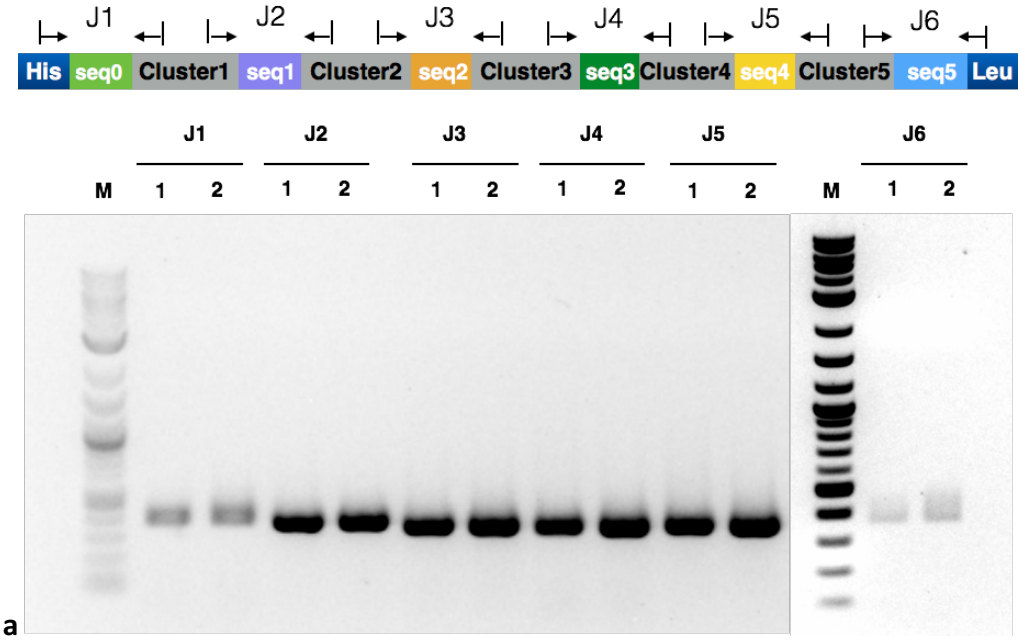


Figure 6.5|Screening and quality control of nocathiacin I construction. a. BSG amplification by PCR. The sizes of the bands are listed in figure 6.4a. **b.** Colony PCR screening of first step BSG assembly with HCKan_O vector. The sizes of the bands are all 100 bp longer than the BSG amplification products because of the additional sequence amplified by the upstream and downstream primers. Two colonies were picked for each construction. The red arrow indicates the band with wrong size. **c.** BsaI digestion verification for POT assembly. After digestion, the size of backbone POT vector is around 5Kb and size of the second bands should be the size of BSG plus a corresponding promoter and terminator (around additional 700 bp). **d.** Digestion verification by different restriction enzymes for pYFASS-multi-TU cluster assembly. Digestion simulation is on the left panel. **e.** Linearization of pYFASS-multi-TU clusters and YAC10 vector for *in vivo* yeast assembly. The pYFASS-clusters were digested by BsmBI and the gel simulation is on the left panel; the YAC10 accept vector was digested by XbaI, SacI and BamHI and the gel simulation is on the right panel.

After being co-transformed with the fragments, the yeast cells were selected on SC-His-Leu 5-FOA plates. To obtain the correct nocathiacin I pathway YAC assembly, a

junction yeast colony PCR test was designed as an easy primary screening method to select colonies for further verification. Primers were designed to bind upstream and downstream of the seam between the YAC10 arms and each of the multi-TU clusters. For the nocathiacin I pathway, there are six junctions and each junction is around 400bp (Figure 6.6a). Eight colonies were picked and tested with the junction PCR. Two clones showed the right signals for all the junctions and the rest were only partially assembled (Figure 6.6a). To further confirm that all BSGs within the clusters were present, a whole BSG colony PCR was performed for the candidates. The yeast colony PCR result shows that all BSGs exist (Figure 6.6b). As a final confirmation, the whole pathway was sent for sequencing to make sure that no mutation was generated during the assembly process. As the whole YAC-nocathiacin I is more than 60Kb and the linear DNA is not easy to recover in *E. coli*, the pathway was divided into 9 fragments to be amplified by KAPA HF PCR. The PCR products were sent for sequencing. The sequencing result confirms that there is no mutation in strain LWy218 and the strain was forwarded for functional characterization (Figure 6.6c).



a

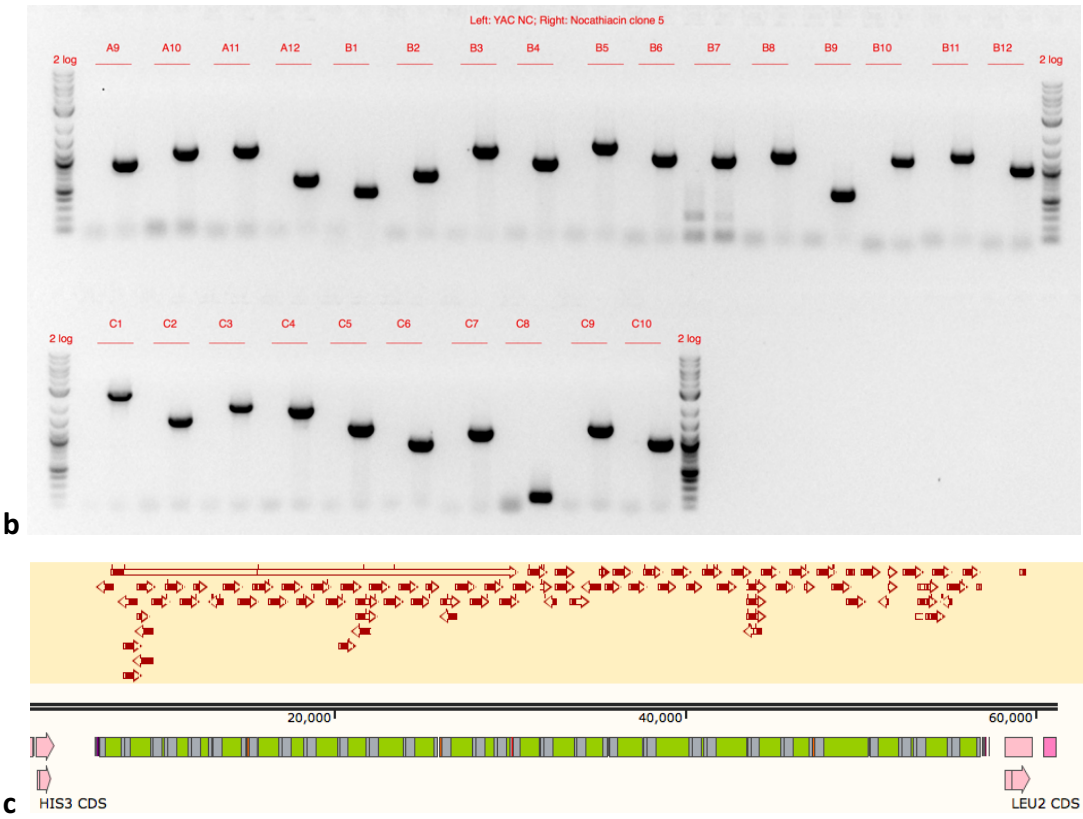


Figure 6.6|Quality control of yeast *in vivo* assembly of the nocathiacin I pathway. **a.** Junction PCR result for primary assembly screening. The primers were designed upstream and downstream of the seam between the clusters and the PCR amplification products are all around 400bp for the six junctions. Two out eight clones have complete junction PCR signals. **b.**BSG colony PCR as a further test of the candidate clones. BY4741 strain with empty YAC10 vector was used as negative control. The result shows that all BSG signals were positive for the YAC-Nocathiacin I clone and negative for the negative control **c.** Sequencing confirmation of the YAC-nocathiacin I assembly. A PCR amplification method was used to sequence the whole pathway and no mutation in BSG was observed in strain LWy218.

6.2.3 Characterization by cell-based assay

After the correct assembly of the nocathiacin I pathway was verified, the expression and antibiotic activity characterization of the pathway was performed. The cell-based overlay assay was done with *B. subtilis*. Before testing the nocathiacin I-containing yeast, a positive control for the overlay assay was established. The purple pigment violacein, which has been introduced in previous chapters, not only is an anti-tumor reagent but also has antibiotic activity against gram-positive pathogens (Choi et al.,

2015). Therefore, the *B. subtilis* overlay assay was tested on violacein producing yeast. A clear inhibition zone was observed surrounding the violacein expressing yeast colony (Figure 6.7). Yeast with YAC10 empty vector was used as a negative control for the assay and no inhibition zone was observed. During the process of pathway diversification by combinatorial assembly, a lot of variants, with or without antibiotic activity, might be generated. To investigate whether the overlay assay can distinguish cells with antibiotic activity from those without in the mixture, the positive control and the negative control were mixed 1:1 for the overlay assay. The result shows that only positive control (PC) have a clear inhibition zone and the negative control (NC) does not, indicating there is no interference between the NC and PC strains and the overlay assay should be a good way for screening pathways with antibiotic activity from a large population of variants (Figure 6.7). After the positive control was set up, the characterization of the nocathiacin I pathway was performed. However, in the *B. subtilis* overlay assay on single colonies of nocathiacin I pathway-carrying yeast, no inhibition zone was observed (Figure 6.8). A further overlay assay on spotted yeast with longer culture of 5 days was done, but still no inhibition zone was generated (Figure 6.8). The strains were also sent to our collaborator, the Wright group, for chemical extraction and analysis, but no signal was detected (data not shown). Therefore, I decided to dig deeper into the transcription and expression profiles of each of the BSGs for troubleshooting.

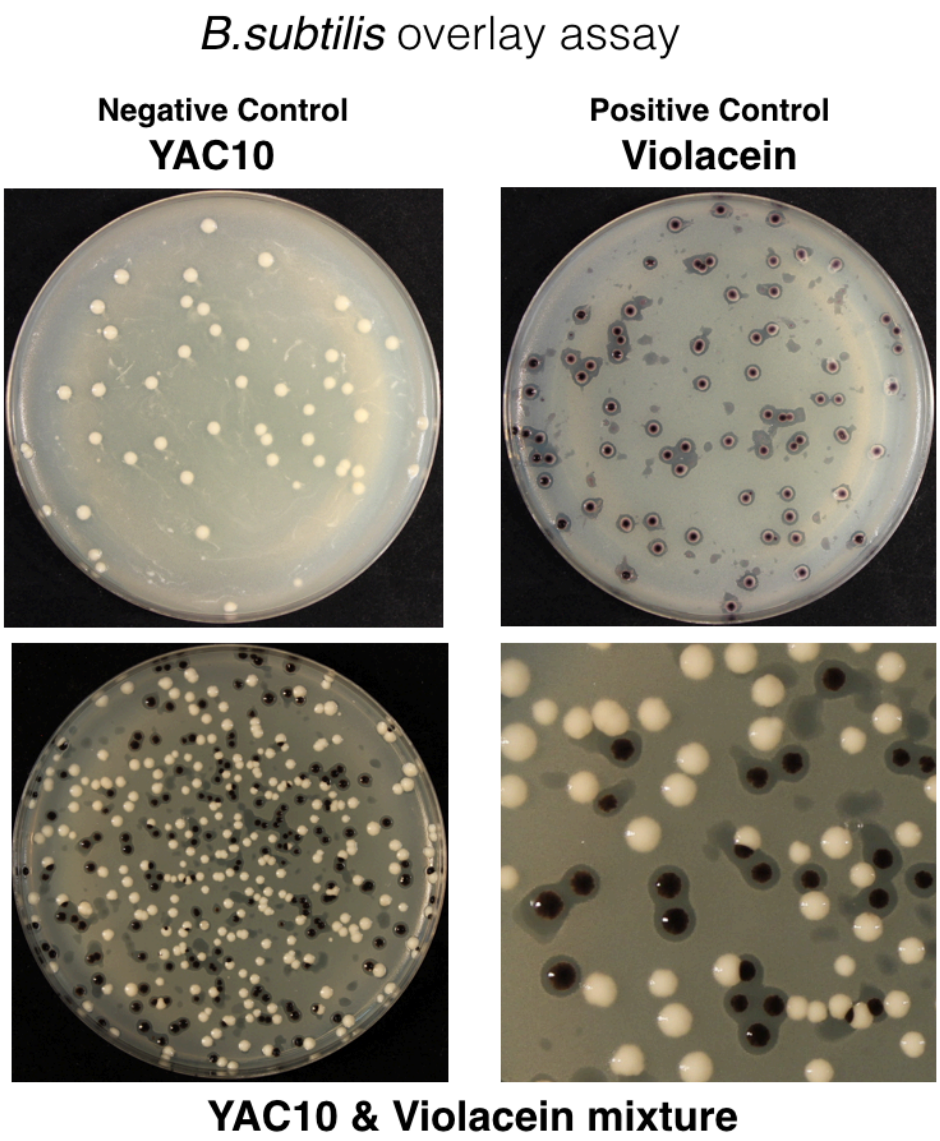


Figure 6.7| *B. subtilis* overlay assay on violacein producing yeast. BY4741 with YAC10 vector was used as negative control; BY4741 strain with violacein pathway was tested with the *B. subtilis* overlay assay and a positive signal with clear inhibition zone was observed. An overlay assay on NC, PC mixture was performed to ensure that the assay can distinguish colonies with antibiotic activity from those without in a cell mixture and the result above confirmed it.

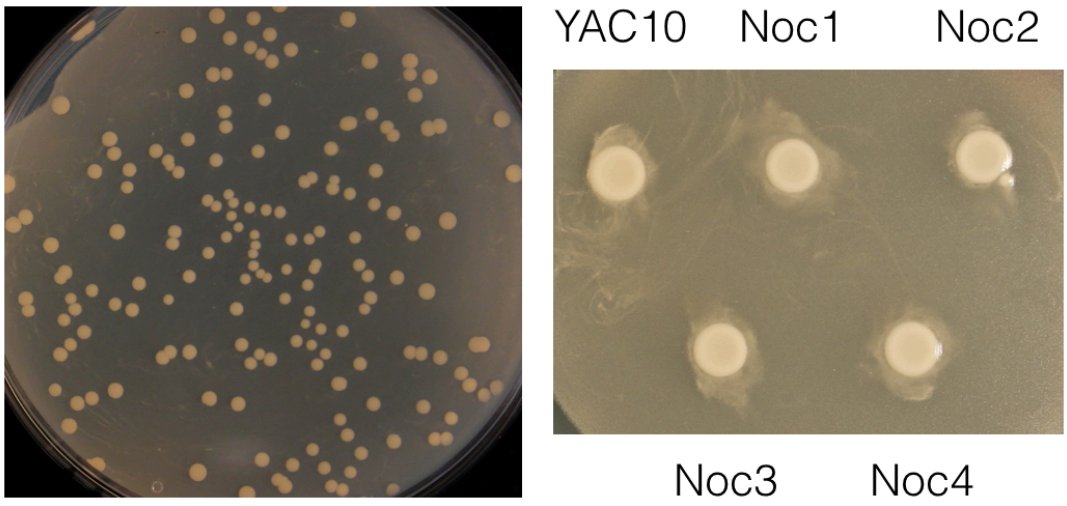


Figure 6.8| *B. subtilis* overlay assay on yeast strain with nocathiacin I pathway. The overlay assay was performed on both single colonies and spotted yeast with the nocathiacin I pathway. No inhibition zone was observed in either case. Yeast with YAC10 was used as a negative control.

6.2.4 Transcriptional characterization by RT-PCR

The first troubleshooting method for the nocathiacin I pathway was to check the transcription status of the BSGs. The whole cell RNA was extracted from the YAC10-nocathiacin I carrying strain LWy218 and reverse transcribed using a commercial kit. The detailed methods of RNA extraction and reverse transcription (RT) are described in Chapter 2.9. Primers were generated for each BSG to amplify a DNA fragment of 100bp-200bp. The housekeeping gene *ACT1* was used as positive control for RT-PCR and the length of the amplification is around 400bp. The RT-PCR shows that except for NocS1 and NocI, all BSGs have positive signals (Figure 6.9). The possibility of low primer amplification efficiency was ruled out by PCR with genome DNA (data not shown) and therefore the transcription status of NocS1 and NocI were not promising. The reason for low transcriptional efficiency was thought to be the promoter. The promoters for NocS1 and NocI are YHR214W_P and YHR128W_P with the measured activities of 6.1 and 4.6 respectively (Y. Guo et al., 2015). However, it is possible that the promoter is not context-independent and somehow the sequence of the BSG makes a difference to transcription efficiency. This is not a big problem since it can be fixed by promoter combinatorial assembly. Before doing the combinatorial

assembly for the two genes, the western blot assay was done to confirm the protein expression of the BSGs.

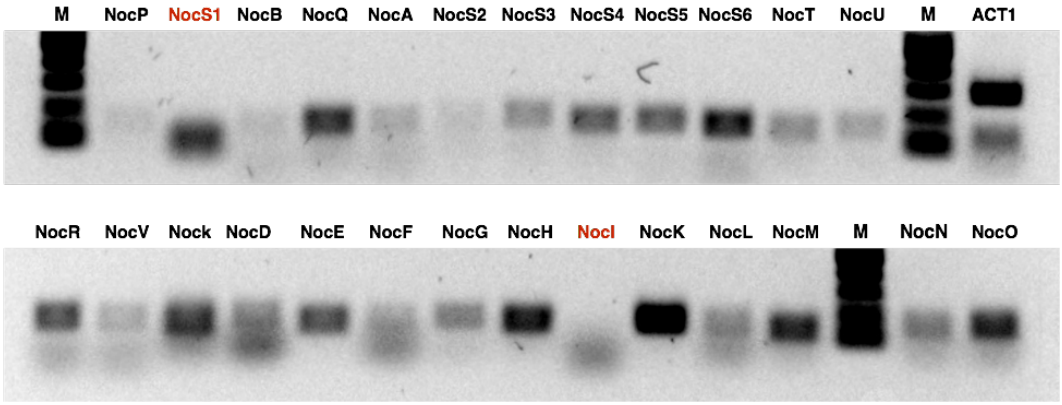


Figure 6.9|RT-PCR result of nocathiacin I BSGs. The RT-PCR signals should be 100-150 bp and were positive for 24 out of 26 BSGs from nocathiacin I pathway. The two with negative signals were NocS1 and NocI that driven by YHR214W_P and YHR128W_P respectively since they only have shorter bands which are more likely to be primer dimer. Housekeeping gene ACT1 was used as positive control and its amplification is around 400 bp.

6.2.5 Protein expression characterization by western blot

As is mentioned in the design section, all BSGs were fused with the FLAG tag (DYKDDDDK) for western blot and only the strongest yeast constitutive promoter P_{TDH3} was used to drive all transcription. The result shows that 23 out of 26 BSGs of nocathiacin I pathway have positive western blot signals (Figure 6.10). The protein sizes of most translated BSGs were correct; only NocF has another lighter band with smaller size, which could indicate degradation of NocF protein (Figure 6.10 and Table 6.4). The three BSGs with negative signals are NocM, NocN and NocO. Unfortunately, these three genes are all of vital importance for producing nocathiacin I: NocM is the precursor peptide; NocO is essential for central core ring formation; NocN is involved in side ring formation and is necessary for decoration of the core scaffold with sugar groups (Table 6.4). Therefore, it is vitally important to figure out the reasons for the negative western blot signal of the three genes. Potential explanations include technical problems of western blot for NocM considering its small peptide size, the codons for these three genes may need to be further optimized to enable proper

translation, or the stability of these proteins may be low in *S. cerevisiae*, etc. A potential solution for translational difficulty is to repeat codon optimization for these genes using some other software such as the "GeneDesign" used for recombinase codon optimization; for solving protein stability problems, on one hand the expression of the three genes can be tuned to avoid instability caused by overexpression, while on the other hand the "N end rules" for estimating protein half-life can be considered for N-terminal engineering of the proteins. However, further experiments could not be conducted due to time limitations.

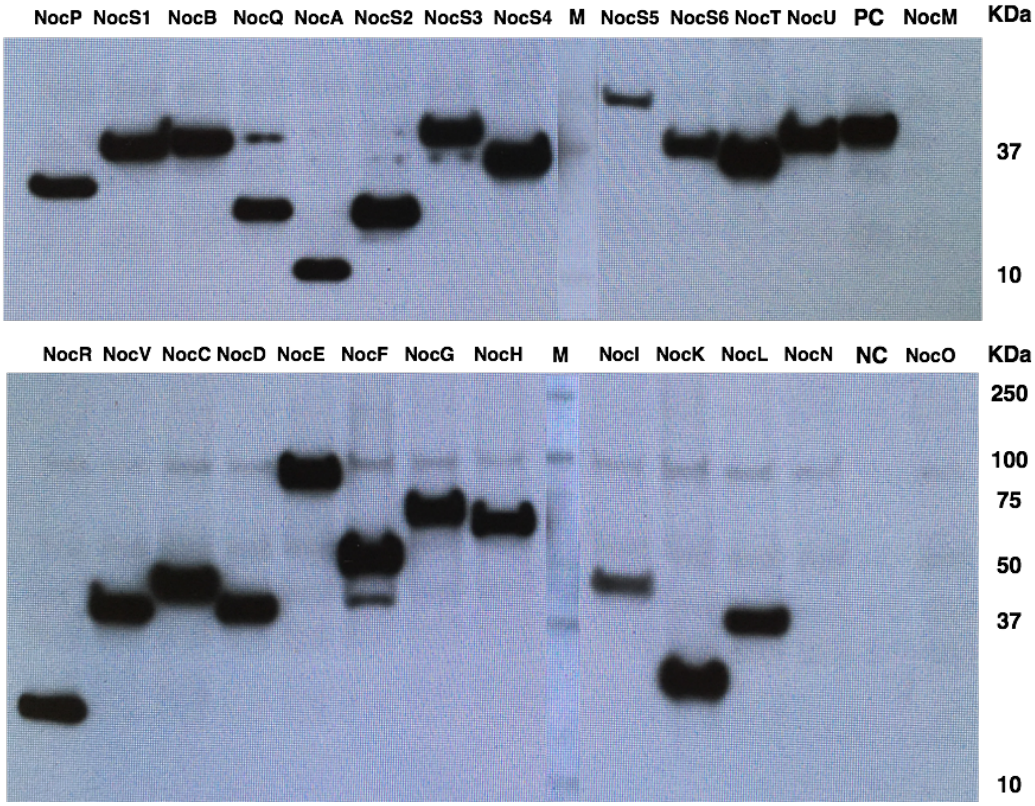


Figure 6.10|Western blot result of nocathiacin I BSGs. The western blot result of the FLAG tagged BSGs shows that 23 out of 26 genes have positive signal and only NocM, NocN and NocO have negative result. The sizes of the proteins with positive signal are all as expected, but NocF has a fainter band with smaller size, which indicates potential degradation. M: Protein ladder (over-exposed during signal developing stage for size reference). PC: Viomycin S FLAG-tagged gene that has been proved to express by colleague Jamie Auxillos. NC: BY4741 with POT2 vector.

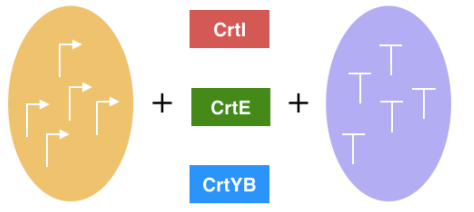
Table 6.4 Size and function information of nocathiacin I BSGs

BSG Name	Protein size(KDa)	Predicted enzyme function (Ding et al., 2010; Hayashi et al., 2014)
NocP	33.78	Regulatory protein
NocS1	42.65	Glycosyltransferase
NocB	43.87	Cytochrome P450 protein
NocQ	23.04	Methyltransferase
NocA	17.1	Hypothetical protein
NocS2	26.48	Sugar N-dimethyltransferase
NocS3	45.78	Sugar 3-C-methyltransferase
NocS4	35.15	NDP-hexose 3-ketoreductase
NocS5	52.48	NDP-hexose 2,3-dehydratase
NocS6	41.36	Sugar aminotransferase
NocT	40.28	Cytochrome P450 protein
NocU	44.2	Cytochrome P450 protein
NocR	20	Glyoxalase/bleomycin resistance protein/Dioxygenase
NocV	42.15	Cytochrome P450 protein
NocC	47.97	Cytochrome P450 protein
NocD	36.97	Dehydratase
NocE	93.37	Dehydratase
NocF	51.19	Dehydrogenase
NocG	69.32	Cyclodehydratase
NocH	62.96	Hypothetical protein (Central core ring formation)
NocI	44.27	ATP-dependent synthetase or ligase
NocK	32.5	Hydrolase/acyltransferase
NocL	43.2	Radical SAM-dependent protein
NocN	47.32	Radical SAM-dependent methyltransferase (Side ring formation)

NocO	34.19	Hypothetical protein (Central core ring formation)
NocM	6.18	Precursor peptide

6.2.6 Proof of principle: combinatorial assembly for β-carotene

To test the efficiency of pathway diversification by combinatorial assembly of promoter and terminator pools, the β-carotene pathway was used for demonstration. Five promoters with strong, medium and weak activity were mixed together and five terminators were mixed together. The promoter mixture and the terminator mixture were used for Golden Gate assembly with the *CrtI*, *CrtE* and *CrtYB* genes. The TUs were further assembled to YeastFab-compatible HO integration vector (pWL121) for final transformation into yeast by frozen yeast transformation method described in section 2.3.4.2.



Promoter Mix		Terminator Mix
Name	Activity	
RPL27A gene	Strong	ACS2 gene
TDH3 gene		ACS1 gene
BCD1 gene	Middle	ENO2 gene
VMA10 gene		CIT1 gene
BRL1 gene	Weak	FUM1 gene

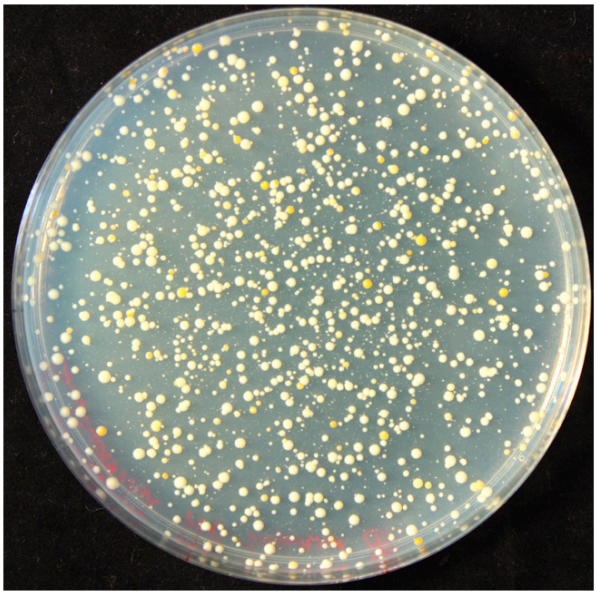


Figure 6.11|β-carotene production diversification by combinatorial assembly. A mixture of 2 strong, 2 medium and 1 weak promoters and a mixture of 5 terminators were used for the combinatorial assembly with *CrtI*, *CrtE* and *CrtYB* gene from the β-carotene pathway. The transformation result shows that there are different yeast colonies with white, yellow and orange color, indicating diversified carotenoid production.

The transformation result of the combinatorial assembly shows that yeast colonies with different β-carotene synthesis circuits have different colors of white, yellow and orange, which indicates the potential for diversifying and optimizing the SynNP pathways with the combinatorial assembly strategy with regulatory elements in future.

6.3 Chapter summary, discussion and future works

In this chapter, the main strategies of applying the synthetic biology platform for engineering natural products pathways in yeast was described. For a pilot application, the platform was specifically designed for screening new antibiotics against tuberculosis. The project is a collaboration between three groups and my contribution to the project includes setting up the workflow for pathway construction from our side, which enables compatibility between combinatorial assembly with YeastFab Golden Gate standards and the YAC yeast *in vivo* assembly standards. I designed and constructed a set of pYFASS vectors and demonstrated the workflow and assembly thoroughly with the 50kb nocathiacin I pathway which contains 26 BSGs. In addition to the DNA assembly method, I contributed a positive control for the cell-based overlay assay with the violacein-producing yeast. Expression characterizations for the RiPPs exemplar pathway nocathiacin I, including overlay assay, RT-PCR and western blot, were performed.

Though the assembled nocathiacin I pathway was not functional in yeast, troubleshooting has narrowed down problems to the expression of the precursor peptide NocM and two other genes involved in core chemical scaffold formation. Future work should be done to fix the expression problems of NocM, NocN and NocO by possible solutions like re-codon optimization and stability optimization for the proteins.

However, even if the expression of all BSGs is confirmed, the normal function of all the BSGs in yeast can hardly be guaranteed. Protein engineering of enzymes with poor function can be attempted, but it might require great effort and time from researchers to ensure the function of every BSG. Another way is to take advantage of

the gradually enlarging BSG library in the project and construct pathway libraries by combinatorial assembly of the enzymes. Constructing and screening effective pathways from the library with so many pathway classes and huge combinatorial possibilities is also incredibly challenging, therefore it is necessary to establish a fully automated assembly platform for high-throughput pathway construction and develop automated and more sensitive assays for antibiotic activity testing and screening. Specific biosensors for target products or intermediate chemicals should be developed for both monitoring metabolic fluxes of the pathways and selection for pathways with higher production.

The combinatorial assembly of regulatory elements like promoters and terminators has been proven effective in pathway optimization with the fact that the promoters and terminators have been characterized. Therefore, combinatorial assembly of BSGs should be more carefully considered. In my opinion, plans on standardized BSG function characterization assays should be added and agreed within the consortium. It is risky to blindly use uncharacterized parts for combinatorial assembly and it can be a waste of time and money to do large-scale screening from libraries composed of uncharacterized genes. Good function quantifications of the BSGs will enable better and more rational design of the pathways to work properly in heterologous hosts and make it more efficient to engineer the pathways for broader applications.

Chapter 7 Summary and future perspectives

My PhD project mainly focused on the characterization and application of tyrosine recombinases. To serve the purpose of controllable co-SCRaMbLE of the normal synthetic chromosomes and the tRNA Neochromosome for the Sc 2.0 consortium, two orthogonal recombination activation systems, CrePBD and DreEBD, were constructed and characterized in *S. cerevisiae*. Primary SCRaMbLE assays by Cre have been performed on SynII, SynIXR, SynV and some other synthetic chromosomes like SynII-XII and SynIII-VI-IX. The characterization of SCRaMbLE by Dre on the tRNA Neochromosome will be carried out in the near future by other lab members. Besides being used as a SCRaMbLE tool for accelerated evolution of the synthetic yeast, the orthogonal recombination systems can also be applied to build logic gates in yeast. Though not all SSR-LBD fusion proteins have been properly induced to activate recombination, further engineering of the proteins by linker optimization or error-prone PCR is likely to enable the expected inducibility of other SSR-LBDs, like Vika, VCre and SCre, and will also expand the catalogue of bioparts available for designing and constructing more eukaryotic computational devices and synthetic chromosomes in future.

Many aspects can be investigated through SCRaMbLE of the synthetic yeast, including evolving a minimal yeast cell, and evolving strains tolerant to higher temperature or higher ethanol concentration and many other industrial conditions. In my doctoral study, a special interest was the effect of SCRaMbLE on heterologous expression of pathways of interest in the synthetic yeast and chassis optimization. A "SCRaMbLE-in" device was designed to integrate a pathway of interest randomly into the $loxP_{sym}$ sites distributed on the synthetic chromosomes and at the same time drive genome variation by Cre recombination. On one hand, random genome integration can optimize the integration locus and genome context for the pathway; on the other, random genome rearrangement provides the chance that the metabolism fluxes may be altered and adjusted for more balanced resource distribution with less burden on the host. The *SynII* strain and two pigment synthesis pathways, violacein and β-carotene, were used for demonstration. With *in vivo* evolution by SCRaMbLE-in and

continuous SCRaMbLE, the yield of violacein was increased by almost 17-fold compared to the HO locus-integrated pathway, proving the potential of the synthetic yeast platform for metabolic engineering by chassis optimization. In the near future, the SCRaMbLE-in device can be used in strains with more synthetic chromosomes and with joint SCRaMbLE of the tRNA Neochromosome. More genome variations will be generated to result in mutant strains with further diversified phenotype that can be screened for different purposes. The synthetic yeast and the SCRaMbLE-in device provide a universal platform and method for metabolic engineering. Besides the demonstration pathways, its application can be expanded to any pathway for stable and balanced heterologous expression with yeast as host. Nonetheless, there are a few challenges which must be met for broader applications. For the purpose of selecting yeast variants with the highest pathway production, an automated and more sensitive selection method is necessary, especially for pathways with colorless and odorless intermediates or products. One solution could be developing biosensors for the target chemicals and designing suitable riboswitches to facilitate automated screening, by FACS for example. Another challenge is to identify relevant genome variation(s) that contribute to the phenotype for the SCRaMbLEed mutants. Since the SCRaMbLE process is completely random, many genome rearrangements can be generated but not all of them are related to the change of phenotype. The methods described in this thesis, finding shared phenotype and genotype among the variants and testing with gene complementation experiments, provide basic strategies to narrow down the modified genotype. However, it is not guaranteed that shared phenotype and genotype will always be captured and the chance is even lower if the phenotype results from comprehensive impacts from multiple genome variations. Therefore, systems biology analysis, computational methods and multi-omics approaches (Chae, Choi, Kim, Ko, & Lee, 2017) must be developed for analyzing the genotype-phenotype relationship, which will help us to better understand the interaction between the heterologous pathway and the local metabolism to facilitate future rational design of novel artificial chromosomes. There are also some technical issues to be pointed out. One problem to be further optimized is the off-target

integration effect potentially caused either by faulty site recognition of Cre or by yeast homologous recombination. The backbone of the SCRaMbLE-in device and the recombinase expression device are both based on the pRS41X centromeric vectors, and there might be interactions between the auxotrophic marker gene sequence in the vectors and genome. Backbone sequence optimization can be performed to avoid off-target integration. Another factor affecting the efficiency of SCRaMbLE-in is the loss rate of Ura3 containing plasmids during counter-selection, since if the *URA3*-containing marker is not lost, even the pathway-integrated strains will not survive and the integration efficiency can be therefore compromised. A modified version of the SCRaMbLE-in device was designed to eliminate *URA3* expression after Cre-mediated recombination by placing the $loxP_{sym}$ sites between the promoter and ORF of the Ura3 expression unit (Figure 7.1). If excision happens between the two $loxP_{sym}$ sites, the promoter and CDS of the *URA3* expression cassette will be separated, thereby stopping the transcription of the Ura3 gene; if inversion happens, the second promoter on the other side of the $loxP_{sym}$ will still enable the transcription of Ura3 and cells containing such variants will not survive with 5-FOA counter selection. Functional characterization of the revised device should be performed in future.

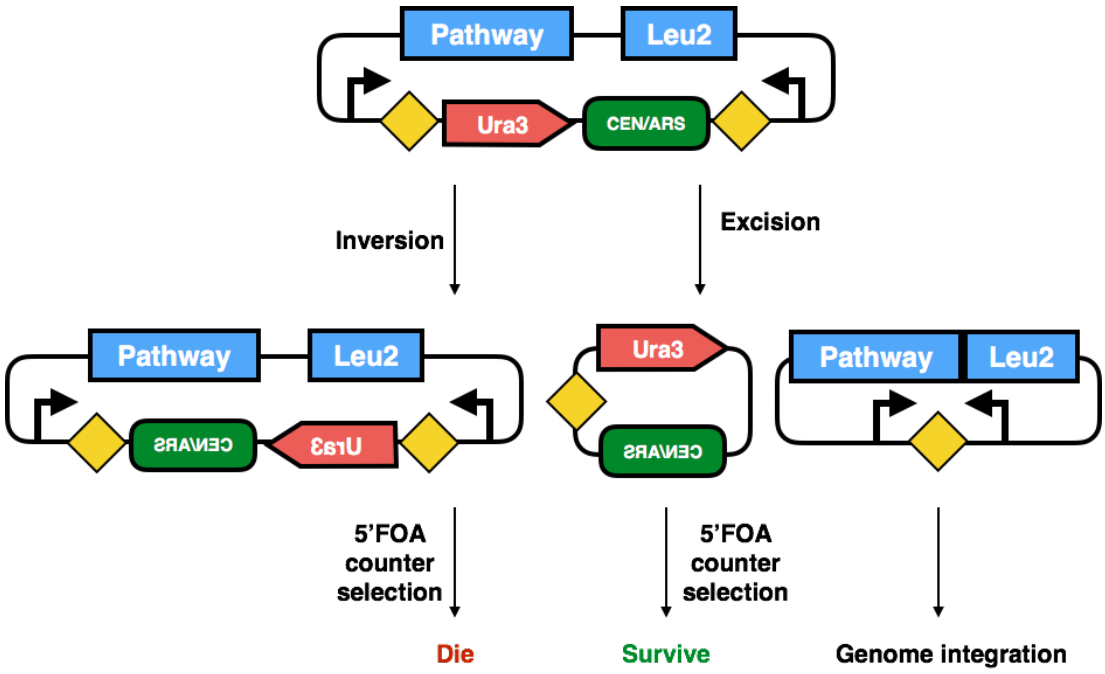


Figure 7.1| Modified SCRaMbLE-in device for higher integration efficiency.

Besides the *in vivo* SCRaMbLE study, the recombinases were also applied *in vitro* for pathway diversification. Upon the successful expression and purification of Cre, Dre and VCre, the recombination orthogonality, excision efficiency and integration efficiency were tested and quantified. The degree of recombination *in vitro* is not 100%, with excision rate around 30% to 40% and integration rate only around 0.5%. In order to improve the integration rate, an LE/RE site mutation strategy was applied. A set of previously generated LE/RE mutant sites that have been tested *in vivo* were chosen for test *in vitro* and their excision and integration rates were quantified with the same strategy. The integration efficiency was improved 3-fold with a loxJT15-loxJTZ17 mutant pair, while its integration efficiency tested in *E. coli* was 1500-fold that of the WT loxP pair (Thomson et al., 2003b). This indicates that the *in vitro* integration process is quite different from that *in vivo*. The recombination conditions should be further optimized for the new recombinases. Moreover, a new and larger recombination site mutant library should be designed and screened *in vitro* for the selection of pairs with higher integration efficiency. Serine integrases are also worth trying with the design for metabolic engineering described here as an alternative for the tyrosine recombinases. If in future the integration efficiency could be raised sufficiently, multiple gene regulation by promoter integration can be achieved without additional *in vivo* selection by multiple auxotrophic markers. The *in vitro* recombinase integration toolkit provides an alternative way to regulate the key genes in a pathway by simple one-pot integration instead of repetitive library construction. In addition to the integration of promoters, the element of interest can be any piece of DNA, like the coding sequences of different genes, transcription units, multi-gene clusters, enhancers, replication origins, etc. Compared with *in vivo* recombination, *in vitro* recombination is faster and easier to control, which makes the recombinases good candidates to build cell-free gene networks and sensors for broad applications. Following the work on developing recombinase tools for metabolic engineering, the SynNP collaborative project provides a platform to express and engineer novel synthetic pathways in yeast for the discovery of new antibiotics against tuberculosis. On top of the two demonstration pathways, violacein and β-carotene, I further

worked on more pathway types including the Nocathiacin I pathway from the RiPPs class. A modular, hierarchical assembly workflow with alternative BsaI and BsmBI Golden Gate assembly and yeast *in vivo* assembly was developed to flexibly construct pathways of theoretically any size and gene numbers. This also enables combinatorial assembly of the promoter, terminator and BSG libraries. With the standardized workflow, the 50Kb Nocathiacin I pathway with 29 TUs was assembled to the YAC vector in yeast. Though the pathway has not shown antibiotic activity against the two gram-positive strains used in testing, the on-going transcription and translation examination by RT-PCR and western blot narrowed down the problem to the expression of three key genes in the pathway, and further optimization and engineering on these genes can be carried out to deal with the problem. In addition to the strategy of combinatorial assembly, the recombinase-based *in vivo* and *in vitro* toolkit can also be applied for optimizing antibiotic production by pathway and chassis level evolution. For example, the *in vitro* recombination system can be used to switch tailoring enzyme expression units or BSG clusters for the purpose of chemical diversification. As to the screening strategy, the cell-based overlay assay currently serves well for selecting variants with antibiotic activity regardless of whether the target chemical has color or odor. However, one condition for the overlay assay is that the yield of the antibiotic has to reach the MIC (minimal inhibition concentration) and those with yield below the MIC will not be detected. Therefore, more sensitive methods for detecting yeast variants with negative if not lethal effect on the test pathogens are necessary for the initial screening of new antibiotics. Subsequent optimization can be performed to further improve the production for full antibiotic function and future industrial fermentation and commercialization.

Appendices

Appendix I DNA sequences

Gene name: Cre
Description: For expression in *S. cerevisiae* and in *E. coli*, clone from Lindstrom DL's work(Lindstrom & Gottschling, 2009)
Sequence:
ATGTCCAATTTACTGACCGTACACCAAAATTTGCCTGCATTACCGGTCGATGCAACGAGTGATGAGGTTCGCA
AGAACCTGATGGACATGTTCAGGGATCGCCAGGCGTTTTCTGAGCATACCTGGAAAATGCTTCTGTCCGTTTG
CCGGTCGTGGGCGGCATGGTGCAAGTTGAATAACCGGAAATGGTTTCCCGCAGAACCTGAAGATGTTCGCGA
TTATCTTCTATATCTTCAGGCGCGCGGTCTGGCAGTAAAAACTATCCAGCAACATTTGGGCCAGCTAAACATGC
TTCATCGTCGGTCCGGGCTGCCACGACCAAGTGACAGCAATGCTGTTTCACTGGTTATGCGGCGGATCCGAAA
AGAAAACGTTGATGCCGGTGAACGTGCAAAACAGGCTCTAGCGTTCGAACGCACTGATTTCGACCAGGTTCGT
TCACTCATGGAAAATAGCGATCGCTGCCAGGATATACGTAATCTGGCATTTCTGGGGATTGCTTATAACACCCT
GTTACGTATAGCCGAAATTGCCAGGATCAGGGTTAAAGATATCTCACGTACTGACGGTGGGAGAATGTTAATC
CATATTGGCAGAACGAAAACGCTGGTTAGCACCGCAGGTGTAGAGAAGGCACTTAGCCTGGGGGTAACTAAA
CTGGTCGAGCGATGGATTTCCGTCTCTGGTGTAGCTGATGATCCGAATAACTACCTGTTTTGCCGGGTCAGAA
AAAATGGTGTTGCCGCGCCATCTGCCACCAGCCAGCTATCAACTCGCGCCCTGGAAGGGATTTTTGAAGCAGC
TCATCGACTGATTTACGGCGCTGAGGATGACTCTGGTCAGAGGTACCTGGCCTGGTCTGGACACAGTGCCCGT
GTCGGAGCCGCGCGAGATATGGCCCGCGCTGGAGTTTCAATACCGGAGATCATGCAAGCTGGTGGCTGGACC
AATGTAAATATTGTCATGAACTATATCCGTAACCTGGATAGTGAAACAGGGGCAATGGTGGTCCTGCTGGAAG
ATGGCGATTGA

Gene name: Dre-yeast
Description: Dre, codon-optimized for expression in *S. cerevisiae*
Sequence:
ATGTCTGAATTGATCATCTCTGGTTCTTCTGGTGGTTTCTTGAGAAACATCGGTAAGGAATACCAAGAAGCTGC
TGAAAACTTCATGAGATTCATGAACGACCAAGGTGCTTACGCTCCAAACACCTTGAGAGACTTGAGATTGGTT
TTCCACTCTTGGGCTAGATGGTGTCACGCTAGACAATTGGCTTGGTTCCCAATCTCTCCAGAAATGGCTAGAGA
ATACTTCTTGCAATTGCACGACGCTGACTTGGCTTCTACCACCATCGACAAGCACTACGCTATGTTGAACATGT
TGTTGTCTCACTGTGGTTTGCCACCATTGTCTGACGACAAGTCTGTTTCTTTGGCTATGAGAAGAATCAGAAGA
GAAGCTGCTACCGAAAAGGGTGAAAGAACCGGTCAAGCTATCCCATTGAGATGGGACGACTTGAAGTTGTTG
GACGTTTTGTTGTCTAGATCTGAAAGATTGGTTGACTTGAGAAACAGAGCTTTCTTGTTCGTTGCTTACAACAC
CTTGATGAGAATGTCTGAAATCTCTAGAATCAGAGTTGGTGACTTGGACCAAACCGGTGACACCGTTACCTTG
CACATCTCTCACACCAAGACCATCACCACCGCTGCTGGTTTGGACAAGGTTTTGTCTAGAAGAACCACCGCTGT
TTTGAACGACTGGTTGGACGTTTCTGGTTTGAGAGAACACCCAGACGCTGTTTTGTTCCCACCAATCCACAGAT
CTAACAAGGCTAGAATCACCACCACCCCATTGACCGCTCCAGCTATGGAAAAGATCTTCTCTGACGCTTGGGTT
TTGTTGAACAAGAGAGACGCTACCCCAAACAAGGGTAGATACAGAACCTGGACCGGTCACTCTGCTAGAGTT
GGTGCTGCTATCGACATGGCTGAAAAGCAAGTTTCTATGGTTGAAATCATGCAAGAAGGTACCTGGAAGAAG
CCAGAAACCTTGATGAGATACTTGAGAAGAGGTGGTGTTTCTGTTGGTGCTAACTCTAGATTGATGGACTCTT
GA

Gene name: Dre-ori
Description: Dre, original sequence from phage, for *E. coli* expression and purification
Sequence:
ATGGgcAGTGAATTAATTATCTCTGGCAGTTCCGGTGGTTTTCTGCGCAACATTGGCAAAGAGTACCAGGAAG
CCGCAGAAAACTTTATGCGGTTCATGAATGATCAGGGTGCTTATGCGCCGAATACTCTGCGCGACCTCCGGTT
GGTGTTTCATTCCTGGGCGCGCTGGTGTCACGCTCGCCAGCTTGCCTGGTTTCCGATCTCACCAGAAATGGCCC
GCGAGTATTTTCTTCAGTTGCATGATGCCGATCTGGCTTCGACCACCATTGATAAGCACTACGCCATGCTTAAC
ATGCTGCTTTCTCATTGTGGCCTTCCGCCACTTTCGGATGATAAAAGTGTTTCTCTGGCTATGCGGCGTATCCG
GCGCGAAGCGGCAACGGAAAAAGGCGAACGAACAGGCCAGGCTATACCGCTGCGATGGGACGACCTGAAAC
TGCTCGACGTCCTGTTGTCCAGGTCAGAACGGCTGGTGGACCTGCGCAACCGGGCTTTTCTCTTTGTTGCATAC
AATACGCTGATGCGTATGTCGGAAATCTCGCGTATTCGTGTAGGAGATCTGGACCAAACAGGTGACACTGTCA

CGCTACATATTTCACACACGAAGACAATAACGACCGCCGCCGGGCTTGATAAGGTGCTTTCCCGTCGCACTACT
GCTGTGCTGAATGACTGGCTGGATGTTTCCGGGCTTCGTGAACATCCTGACGCGGTGCTGTTCCCGCCGATAC
ACCGTAGCAATAAGGCCAGGATTACGACAACGCCCCTTACTGCACCTGCAATGGAGAAAATATTCAGCGACGC
CTGGGTGTTGCTGAATAAAAGGGATGCCACGCCAAACAAAGGGAGATACCGGACGTGGACCGGGCATAGTG
CTCGTGTCGGGGCCGCTATCGATATGGCTGAAAAGCAGGTGTCTATGGTGGAGATTATGCAGGAAGGAACCT
GGAAGAAACCGGAAACACTTATGCGGTATCTGCGCCGTGGCGGTGTGTCTGTGGGGGCTAATAGCCGCTTGA
TGGATTCACTCGAGTAG

Gene name: Vika-yeast
Description: Vika, codon optimized for expression in *S. cerevisiae*
Sequence:
ATGACCGACTTGACCCCATTCCCACCATTGGAACACTTGGAACCAGACGAATTCGCTGACTTGGTTAGAAAGG
CTATCAAGCGCGACCCACAAGCTGGTGCTCACCCAGCTATCCAATCTGCTATCTCTCACTTCCAAGACGAATTC
GTTAGAAGACAAGGTGAATGGCAACCAGCTACCTTGCAAAGATTGAGAAACGCTTGGAACGTTTTCGTTAGAT
GGTGTACCCACCAAGGTATCCCAGCTTTGCCAGCTAGACACCAAGACGTTGAAAGATACTTGATCGAAAGAAG
AAACGAATTGCACAGAAACACCTTGAAGGTTCACTTGTGGGCTATCGGTAAGACCCACGTTATCTCTGGTTTG
CCAAACCCATGTGCTCACAGATACGTTAAGGCTCAAATGGCTCAAATCACCCACCAAAAGGTTAGAGAAAGAG
AAAGAATCGAACAAGCTCCAGCTTTCAGAGAATCTGACTTGGACAGATTGACCGAATTGTGGTCTGCTACCAG
ATCTGTTACCCAACAAAGAGACTTGATGATCGTTTCTTTGGCTTACGAAACCTTGTTGAGAAAGAACAACTTGG
AACAAATGAAGGTTGGTGACATCGAATTCTGTCAAGACGGTTCTGCTTTGATCACCATCCCATTCTCTAAGACC
AACCACTCTGGTCGCGACGACGTTAGATGGATCTCTCCACAAGTTGCTAACCAAGTTCACGCTTACTTGCAATT
GCCAAACATCGACGCTGACCCACAATGTTTCTTGTTGCAAAGAGTTAAGAGATCTGGTAAGGCTTTGAACCCA
GAATCTCACAACACCTTGAACGGTCACCACCCAGTTTCTGAAAAGTTGATCTCTAGAGTTTTCGAAAGAGCTTG
GAGAGCTTTGAACCACGAAACCGGTCCAAGATACACCGGTCACTCTGCTAGAGTTGGTGCTGCTCAAGACTTG
TTGCAAGAAGGTTACTCTACCTTGCAAGTTATGCAAGCTGGTGGTTGGTCTTCTGAAAAGATGGTTTTGAGAT
ACGGTAGACACTTGCACGCTCACACCTCTGCTATGGCTCAAAAGAGAAGACAAAGATGA
Gene name: Vika-ori
Description: Vika, original sequence from phage, for *E. coli* expression and purification
Sequence:
ATGGgcACGGATTTGACCCCCTTCCCACCGCTCGAACATCTGGAGCCCGACGAGTTTGCCGATCTGGTGCGCA
AGGCGATCAAACGTGATCCTCAAGCGGGCGCTCACCCTGCGATCCAATCCGCCATCAGCCACTTTCAAGACGA
ATTTGTCCGTCGCCAGGGAGAGTGGCAACCCGCGACACTGCAACGCTTGCGTAATGCCTGGAATGTATTTGTC
CGATGGTGCACTCACCAAGGTATCCCCGCGTTGCCGGCGCGTCATCAGGACGTCGAGCGTTATTTGATTGAGC
GCCGCAACGAGTTACACCGTAACACTTTAAAAGTACACCTTTGGGCCATTGGTAAGACGCATGTTATTTCCGG
ACTACCGAACCCGTGTGCTCATCGTTATGTCAAAGCGCAAATGGCACAAATCACTCATCAGAAAGTGCGCGAA
CGCGAGCGCATTGAGCAAGCCCCTGCCTTTCGTGAGTCGGATCTGGATCGCCTAACCGAACTTTGGTCTGCGA
CTCGTTCCGTGACCCAACAACGCGACCTCATGATCGTCAGCCTCGCGTACGAAACTCTGCTCCGAAAAAACAA
CCTGGAGCAAATGAAAGTCGGTGACATCGAATTTTGCCAAGACGGCAGCGCACTGATCACCATTCCTTTCAGC
AAGACCAATCACTCGGGCCGCGATGATGTGCGCTGGATCTCCCCCCAAGTCGCCAACCAAGTGCACGCGTATT
TGCAATTGCCCAACATTGATGCCGATCCGCAGTGTTTCTTACTTCAACGGGTCAAGCGCTCAGGAAAGGCGCT
TAACCCAGAGTCACATAACACACTCAATGGTCATCACCCGGTGTCCGAAAAGCTGATCAGCCGGGTGTTTGAA
CGCGCTTGGAGAGCCCTCAACCACGAGACAGGCCCACGCTACACGGGCCATAGCGCACGTGTTGGCGCCGCA
CAAGATTTGTTGCAAGAAGGGTATTCGACCCTCCAAGTGATGCAAGCTGGAGGGTGGAGCTCAGAAAAAATG
GTCCTAAGGTACGGTCGACACCTGCATGCCCATACCAGCGCAATGGCACAAAAACGCCGTCAGCGCCTCGAGT
AG

Gene name: VCre-yeast
Description: VCre, codon optimized for expression in *S. cerevisiae*
Sequence:
ATGATCGAAAACCAATTGTCTTTGTTGGGTGACTTCTCTGGTGTTAGACCAGACGACGTTAAGACCGCTATCCA
AGCTGCTCAAAAGAAGGGTATCAACGTTGCTGAAAACGAACAATTCAAGGCTGCTTTCGAACACTTGTTGAAC
GAATTCAAGAAGAGAGAAGAAAGATACTCTCCAAACACCTTGAGAAGATTGGAATCTGCTTGGACCTGTTTCG
TTGACTGGTGTTTGGCTAACCACAGACACTCTTTGCCAGCTACCCCAGACACCGTTGAAGCTTTCTTCATCGAA
AGAGCTGAAGAATTGCACAGAAACACCTTGTCTGTTTACAGATGGGCTATCTCTAGAGTTCACAGAGTTGCTG
GTTGTCCAGACCCATGTTTGGACATCTACGTTGAAGACAGATTGAAGGCTATCGCTAGAAAGAAGGTTAGAG

AAGGTGAAGCTGTTAAGCAAGCTTCTCCATTCAACGAACAACACTTGTTGAAGTTGACCTCTTTGTGGTACAGA
TCTGACAAGTTGTTGTTGAGAAGAAACTTGGCTTTGTTGGCTGTTGCTTACGAATCTATGTTGAGAGCTTCTGA
ATTGGCTAACATCAGAGTTTCTGACATGGAATTGGCTGGTGACGGTACCGCTATCTTGACCATCCCAATCACCA
AGACCAACCACTCTGGTGAACCAGACACCTGTATCTTGTCTCAAGACGTTGTTTCTTTGTTGATGGACTACACC
GAAGCTGGTAAGTTGGACATGTCTTCTGACGGTTTCTTGTTCGTTGGTGTTTCTAAGCACAACACCTGTATCAA
GCCAAAGAAGGACAAGCAAACCGGTGAAGTTTTGCACAAGCCAATCACCACCAAGACCGTTGAAGGTGTTTT
CTACTCTGCTTGGGAAACCTTGGACTTGGGTAGACAAGGTGTTAAGCCATTCACCGCTCACTCTGCTAGAGTT
GGTGCTGCTCAAGACTTGTTGAAGAAGGGTTACAACACCTTGCAAATCCAACAATCTGGTAGATGGTCTTCTG
GTGCTATGGTTGCTAGATACGGTAGAGCTATCTTGGCTCGCGACGGTGCTATGGCTCACTCTAGAGTTAAGAC
CAGATCTGCTCCAATGCAATGGGGTAAGGACGAAAAGGACTGA

Gene name: VCre-ori
Description: VCre, original sequence from phage, for *E. coli* expression and purification
Sequence:
ATGGgcATAGAAAACCAACTCTCCCTGCTCGGTGACTTTTCCGGAGTTAGGCCGGATGATGTCAAAACTGCGA
TCCAAGCGGCCCAGAAAAAAGGCATCAATGTTGCTGAGAACGAGCAGTTCAAGGCAGCATTTGAGCACCTGC
TTAATGAGTTCAAAAAGCGGGAAGAGCGTTACTCTCCCAACACGCTGCGCCGGCTGGAAAGCGCGTGGACCT
GTTTTGTTGACTGGTGTTTGGCGAATCATCGCCATTCTTTGCCGGCGACACCGGATACCGTAGAAGCCTTCTTT
ATTGAGCGAGCGGAAGAGTTACACCGAAACACGTTATCGGTGTATCGCTGGGCGATCTCCCGTGTTCACCGTG
TGGCTGGCTGTCCGGATCCCTGCCTCGATATCTACGTGGAAGACCGGCTGAAAGCCATTGCTCGCAAGAAAGT
GAGAGAGGGCGAGGCCGTGAAACAGGCGTCCCCGTTCAATGAACAACACCTGCTTAAACTGACCTCGCTTTG
GTACCGCAGTGACAAGTTACTGCTTCGCCGCAATCTGGCATTGCTGGCCGTGGCTTATGAGTCAATGCTGCGA
GCTTCAGAGCTGGCGAATATTCGCGTGAGTGATATGGAGCTGGCCGGCGACGGTACCGCGATACTGACTATC
CCCATCACCAAAACCAATCACAGCGGGGAGCCGGATACCTGCATACTCAGTCAGGATGTGGTCAGCTTGCTGA
TGGATTACACCGAAGCCGGCAAGCTGGATATGAGCAGTGACGGCTTTTTGTTTGTGGGCGTGTCCAAGCACA
ACACCTGCATCAAACCCAAGAAGGACAAGCAGACCGGCGAGGTCCTGCACAAGCCAATCACCACCAAAACGG
TTGAGGGCGTGTTCTATTCTGCATGGGAAACTTTGGACTTAGGACGTCAGGGGGTGAAGCCTTTCACGGCACA
CAGTGCGCGTGTTGGCGCTGCTCAAGATCTGCTTAAAAAAGGCTATAACACCCTGCAGATCCAACAGTCTGGC
CGATGGTCTAGCGGTGCAATGGTGGCGAGATATGGCAGGGCGATACTGGCGCGTGATGGTGCAATGGCTCA
TAGCCGTGTCAAAACACGCTCAGCCCCTATGCAGTGGGGAAAAGATGAAAAGGATCTCGAGTAG

Gene name: SCre-yeast
Description: SCre, codon optimized for expression in *S. cerevisiae*
Sequence:
ATGTCTTTGTTGACCACCAACAACCACTCTGTTGCTTTGTCTTACGGTGAACCACCATCTACCTTGAACGACTCT
TTGAAGGACTCTTACCAAAGATCTACCGACGAATTGCAAGCTTTGTTGTCTAAGCCATTGGCTCAATTGACCGA
CGCTGACAAGTTGAGAATCAGAGAAATCACCCAAGCTAAGTTGAAGCACTTCTTGGACAACGGTCACAGAAC
CAGAAGAGCTAACACCTGGAGAGCTTTGATGTCTAGATGGGCTAAGTTCGAATCTTGGTGTTTGACCAACAAC
TTGACCCCATTGCCAGCTACCCCAGAAGTTGTTGCTACCTTCATCGAATACTACCAAGCTTCTTCTTACACCACC
TTGTCTCAATACGCTTGGGCTATCAACTCTTTCCACGTTGAATGTGGTTTGTTGTCTCCAGTTTCTTCTAAGACC
GTTCAAGACAAGCAAAACGAAATCAGAATCGTTAAGTTGGAATCTGGTGGTTTGGCTCAAGAACAAGCTACCC
CATTCAGATTGCACCACTTGCAAATGTTGATCGAATCTTACGGTGAATCTGAAAGATTGTTGGACAAGAGAAA
CTTGGCTTTGTTGAACATCGCTTACGAATCTTTGTTGAGAGAATCTGAATTGTTGAGAATCAAGGTTGGTCACT
TGAAGTCTACCTTCGAAGGTGACTACGTTTTGTCTGTTCCATACACCAAGACCAACGACTCTGGTGAAGAAGA
AGTTGTTAACATCACCCCATTGGGTTTCAAGTTGATCCAAAGATACATCCAAGGTGCTGGTTTGACCAAGGAA
GACTACTTGTTCCAACCAATCGGTAGATCTAACAAGGTTTCTGTTCAAGCTAAGCCAATGTCTACCAGAACCGT
TGACAGAGTTTTCTTGTGGGCTTTCGAATCTTTGGGTATCGACAGACACTCTGCTTGGTCTGGTCACTCTGCTA
GAATCGGTGCTGCTCAAGACTTGTTGGCTGCTGGTTACTCTATCGCTCAAATCCAAGAAAACGGTAGATGGAA
GTCTCCAATGATGGTTTTGAGATACGGTAAGGACATCAAGGCTAAGGAATCTGCTATGGCTAAGATGTTGGCT
GAAAGAAGATGA
Gene name: SCre-*E. coli*
Description: SCre, codon optimized for expression and purification in *E. coli*
Sequence:
ATGGgcGTTTCTCTGCTGACCACCAACAACCACTCTGTTGCTCTGTCTTACGGTGAACCGCCGTCTACCCTGAAC
GACTCTCTGAAAGACTCTTACCAGCGTTCTACCGACGAACTGCAGGCTCTGCTGTCTAAACCGCTGGCTCAGCT

GACCGACGCTGACAAACTGCGTATCCGTGAAATCACCCAGGCTAAACTGAAACACTTCCTGGACAACGGTCAC
CGTACCCGTCGTGCTAACACCTGGCGTGCTCTGATGTCTCGTTGGGCTAAATTCGAATCTTGGTGCCTGACCAA
CAACCTGACCCCGCTGCCGGCTACCCCGGAAGTTGTTGCTACCTTCATCGAATACTACCAGGCTTCTTCTTACAC
CACCCTGTCTCAGTACGCTTGGGCTATCAACTCTTTCCACGTTGAATGCGGTCTGCTGTCTCCGGTTTCTTCTAA
AACCGTTCAGGACAAACAGAACGAAATCCGTATCGTTAAACTGGAATCTGGTGGTCTGGCTCAGGAACAGGC
TACCCCGTTCCGTCTGCACCACCTGCAGATGCTGATCGAATCTTACGGTGAATCTGAACGTCTGCTGGACAAAC
GTAACCTGGCTCTGCTGAACATCGCTTACGAATCTCTGCTGCGTGAATCTGAACTGCTGCGTATCAAAGTTGGT
CACCTGAAATCTACCTTCGAAGGTGACTACGTTCTGTCTGTTCCGTACACCAAAACCAACGACTCTGGTGAAGA
AGAAGTTGTTAACATCACCCCGCTGGGTTTCAAACTGATCCAGCGTTACATCCAGGGTGCTGGTCTGACCAAA
GAAGACTACCTGTTCCAGCCGATCGGTCGTTCTAACAAAGTTTCTGTTCAGGCTAAACCGATGTCTACCCGTAC
CGTTGACCGTGTTTTCCTGTGGGCTTTCGAATCTCTGGGTATCGACCGTCACTCTGCTTGGTCTGGTCACTCTGC
TCGTATCGGTGCTGCTCAGGACCTGCTGGCTGCTGGTTACTCTATCGCTCAGATCCAGGAAAACGGTCGTTGG
AAATCTCCGATGATGGTTCTGCGTTACGGTAAAGACATCAAAGCTAAAGAATCTGCTATGGCTAAAATGCTGG
CTGAACGTCGTCTCGAGTAG

Gene name: EBD
Description: Estrogen binding domain, expressed in-frame with recombinase
Sequence:
CTCGAGCCATCCGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAACGCTCTAAGG
AGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACT
CTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACA
GGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGG
TCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAACACCCGGG
GAAGCTCCTGTTTGCTCCTAACTTGCTCCTGGACAGGAATCAAGGTAAATGTGTGGAAGGCATGGTGGAGATC
TTCGACATGCTGCTGGTTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA
AATCTATTATTTTGCTTAATTCTGGAGTATACACATTTCTGTCCAGCACCCTGAAGTCTCTCGAGGAGAAGGAC
CATATCCACCGAGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGC
AGCAGCAGCACCAGCGGCTAGCCCAGCTCCTCCTCCTCCTCTCCCACATCAGGCACATGAGTAACGAAGGCAT
GGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTACCCCTCTATGACCTGCTGCTGGAGATGCTAGACGC
CCACCGCCTACATGCGCCCACTAGTCGTGGAGGGGCATCCGTGGAGGAGACGGACCAAAGCCACTTGGCCAC
TGCGGGCTCTACTTCATCGCATTCCTTGCAAAAGTATTACATCACGGGGGAGGCAGAGGGTTTCCCTGCCACA
GTCTGA

Gene name: PBD
Description: Progesterone binding domain, expressed in frame with recombinase
Sequence:
CTCGAGCCATTGGACGCTGTTGCTTTGCCACAACCAGTTGGTGTTCCAAACGAATCTCAAGCTTTGTCTCAAAG
ATTCACCTTCTCTCCAGGTCAAGACATCCAATTGATCCCACCATTGATCAACTTGTTGATGTCTATCGAACCAGA
CGTTATCTACGCTGGTCACGACAACACCAAGCCAGACACCTCTTCTTCTTTGTTGACCTCTTTGAACCAATTGGG
TGAAAGACAATTGTTGTCTGTTGTTAAGTGGTCTAAGTCTTTGCCAGGTTTCAGAAACTTGCACATCGACGACC
AAATCACCTTGATCCAATACTCTTGGATGTCTTTGATGGTTTTCGGTTTGGGTTGGAGATCTTACAAGCACGTTT
CTGGTCAAATGTTGTACTTCGCTCCAGACTTGATCTTGAACGAACAAAGAATGAAGGAATCTTCTTTCTACTCTT
TGTGTTTGACCATGTGGCAAATCCCACAAGAATTCGTTAAGTTGCAAGTTTCTCAAGAAGAATTCTTGTGTATG
AAGGTTTTGTTGTTGTTGAACACCATCCCATTGGAAGGTTTGAGATCTCAAACCCAATTCGAAGAAATGAGATC
TTCTTACATCAGAGAATTGATCAAGGCTATCGGTTTGAGACAAAAGGGTGTTGTTTCTTCTTCTCAAAGATTCT
ACCAATTGACCAAGTTGTTGGACAACTTGCACGACTTGGTTAAGCAATTGCACTTGTACTGTTTGAACACCTTC
ATCCAATCTAGAGCTTTGTCTGTTGAATTCCCAGAAATGATGTCTGAAGTTATCGCTATGCACGCTTACGGTTG
A

Homologous arms sequences in YFPASS vectors

Arm name	Sequence
Seq0	TCAGCGATCTGCGATATATTTATGTCTGAAATTGTCGTGAGAAATATTTACAAGCATCCGTGTGT CTGTTGTAAAAAATAATCAACACTATCCTGGTTCG
Seq1	CTGAATACTATGATTCCTGTGGACGAAAAGATCAAGAGCTTGTAATCTTGGGACCACGTCGTAA GCACCTTACATAATACTTACGAGGACTATTATTTCT
Seq2	GACATGATGTAAAGATATGTCAGTAACTTAAATGGGCTAAACGACATATTGTTCACTAAAGGCC GCATTCGTCATAACCTTTAACAATGACTCTCGAGAT
Seq3	TAGTAAGTGGGTCATTTTCTATAACTCACTTATTCGTACGACTTGGACTAACGTATTCTTTGGTTA ATCAGTCCTTGATGAACATCTGTTGTCCTAGGAT
Seq4	GAAGATTACAGAACTTATATGGTTCACACTTCCTGTCAAGACTGGGAAGTGCACACGTGTCTTG GCGAACACATCTACCATTTACCACCACAAAAGTTGA
Seq5	AAAACTTCTGACATTTATTCTTTTTATACCCAGAGTACTTTGAGACTAATCATTCTATAATTAGTA ACGAAACCAGGCTGGACAGTCCAAGCGTAAAAGG

Appendix II Oligo sequences

II-1 Oligos for Dre synthesis by BAG protocol

Description	Sequence
opDre-ORF.01.o01	GGTCTCAAATGTCTGAATTGATCATCTCTGGTTCTTCTGGTGGTTTCTTGAGAAACATCGGTAAGGAATACCAAGAAGC
opDre-ORF.01.o02	GGAGCGTAAGCACCTTGGTCGTTCATGAATCTCATGAAGTTTTCAGCAGCTTCTTGGTATTCCTTACCGATGTTT
opDre-ORF.01.o03	AACGACCAAGGTGCTTACGCTCCAAACACCTTGAGAGACTTGAGATTGGTTTTCCACTCTTGGGCTAGATGGTGTC
opDre-ORF.01.o04	AAGTATTCTCTAGCCATTTCTGGAGAGATTGGGAACCAAGCCAATTGTCTAGCGTGACACCATCTAGCCCAAGAGTGGA
opDre-ORF.01.o05	CTCTCCAGAAATGGCTAGAGAATACTTCTTGCAATTGCACGACGCTGACTTGGCTTCTACCACCATCGACAAGCAC
opDre-ORF.01.o06	AGACAATGGTGGCAAACCACAGTGAGACAACAACATGTTCAACATAGCGTAGTGCTTGTCGATGGTGGTAGAAGC
opDre-ORF.01.o07	CACTGTGGTTTGCCACCATTGTCTGACGACAAGTCTGTTTCTTTGGCTATGAGAAGAATCAGAAGAGAAGCTGCTACC
opDre-ORF.01.o08	GTCGTCCCATCTCAATGGGATAGCTTGACCGGTTCTTTCACCCTTTTCGGTAGCAGCTTCTCTTCTGATTCTTCT
opDre-ORF.01.o09	GCTATCCCATTGAGATGGGACGACTTGAAGTTGTTGGACGTTTTGTTGTCTAGATCTGAAAGATTGGTTGACTTGAGAA
opDre-ORF.01.o10	TTTCAGACATTCTCATCAAGGTGTTGTAAGCAACGAACAAGAAAGCTCTGTTTCTCAAGTCAACCAATCTTTCAGATCT
opDre-ORF.01.o11	TACAACACCTTGATGAGAATGTCTGAAATCTCTAGAATCAGAGTTGGTGACTTGGACCAAACCGGTGACACCG
opDre-ORF.01.o12	CCAAACCAGCAGCGGTGGTGATGGTCTTGGTGTGAGAGATGTGCAAGGTAACGGTGTCACCGGTTTGGTCCAA
opDre-ORF.01.o13	ATCACCACCGCTGCTGGTTTGGACAAGGTTTTGTCTAGAAGAACCACCGCTGTTTTGAACGACTGGTTGGACGTTTC
opDre-ORF.01.o14	GATCTGTGGATTGGTGGGAACAAAACAGCGTCTGGGTGTTCTCTCAAACCAGAAACGTCCAACCAGTCGTTCAAAAC
opDre-ORF.01.o15	GTTTTGTTCCCACCAATCCACAGATCTAACAAGGCTAGAATCACCACCACCCCATTGACCGCTCCAGCTATGG
opDre-ORF.01.o16	TGGGGTAGCGTCTCTCTTGTTCAACAAAACCCAAGCGTCAGAGAAGATCTTTTCCATAGCTGGAGCGGTCAATGG
opDre-ORF.01.o17	TTGAACAAGAGAGACGCTACCCCAAACAAGGGTAGATACAGAACCTGGACCGGTCACTCTGCTAGAGTTGGTGC
opDre-ORF.01.o18	TACCTTCTTGCATGATTTCAACCATAGAAACTTGCTTTTCAGCCATGTCGATAGCAGCACCAACTCTAGCAGAGTGACC
opDre-ORF.01.o19	TCTATGGTTGAAATCATGCAAGAAGGTACCTGGAAGAAGCCAGAAACCTTGATGAGATACTTGAGAAGAGGTGGTGTTT
opDre-ORF.01.o20	GGTCTCTCTCAAGAGTCCATCAATCTAGAGTTAGCACCAACAGAAACACCACCTCTTCTCAAGTATCTCAT
opDre-Fuse.01.o20	GGTCTCTCGAGAGAGTCCATCAATCTAGAGTTAGCACCAACAGAAACACCACCTCTTCTCAAGTATCTCAT

II-2 Primers for recombinase amplification

Description	Sequence
Dre F (NcoI)	CCATGGATGTCTGAATTGATCATCTCTGGTTC
Dre R (XhoI)	CTCGAGAAGAGTCCATCAATCTAGAGTTAGCAC
Cre (NcoI) F	CCATGGATGTCCAATTTACTGACCGTACACC
Cre (XhoI) R	CTCGAGATCGCCATCTTCCAGC
Vika F (NcoI)	agcgtgGGTCTCaGATGCCATGGgcACCGACTTGACCCCATTCCC
Vika R (XhoI)	gtgctgGGTCTCgGCTACTCGAGTCTTTGTCTTCTCTTTTGAGCC
VCre F (NcoI)	agcgtgGGTCTCaGATGCCATGGgcATCGAAAACCAATTGTCTTTG
VCre R (XhoI)	gtgctgGGTCTCgGCTACTCGAGGTCCTTTTCGTCCTTACCCC
SCre F (NcoI)	agcgtgGGTCTCaGATGCCATGGgcTCTTTGTTGACCACCAACAACC
SCre R (XhoI)	gtgctgGGTCTCgGCTACTCGAGTCTTCTTTCAGCCAACATCTTAGC
Dre (PBD fuse) F	GGTCTCAAATGGGTGCTTCTGAATTGATCATCTCTGGTTCTTCTG
Dre(PBD fuse) R	GGTCTCTCCAAAGCACCAGAAGCAGAGTCCATCAATCTAGAGTTAGC
Dre (+ GD for EBD fuse)	GGTCTCTCGAGGTCACCAGAGTCCATCAATCTAGAGTTAGC
GG Cre Fuse F	GGTCTCAAATGTCCAATTTACTGACCGTACAC
GG Cre Fuse R	GGTCTCTCGAGATCGCCATCTTCCAGC
Vika F	GGTCTCAAATGACCGACTTGACCCCATTC
Vika R Fuse	GGTCTCTCGAGTCTTTGTCTTCTCTTTTGAGCCATAGC
Vika R ORF	GGTCTCTCTCATCTTTGTCTTCTCTTTTGAGCC
VCre F	GGTCTCAAATGATCGAAAACCAATTGTCTTTGTTGG
VCre R Fuse	GGTCTCTCGAGGTCCTTTTCGTCCTTACCCCATTG
VCre R ORF	GGTCTCTCTCAGTCCTTTTCGTCCTTACCC
SCre F	GGTCTCAAATGTCTTTGTTGACCACCAACAAC
SCre R Fuse	GGTCTCTCGAGTCTTCTTTCAGCCAACATCTTAGCC
SCre R ORF	GGTCTCTCTCATCTTCTTTCAGCCAACATCTTAG

II-3 Primers for promoter and terminator amplification for recombinase TU assembly

Description	Sequence
B-SCW11-B F	GGTCTCACAGTTTAATTACTTGTGTCTTGACGCAATGAG
B-SCW11-B R	GGTCTCTCATTGTGTACAGGATGATGGTTTTCTATTAG
B-TDH3-B F	GGTCTCACAGTACAGTTTATTCCTGGCATCCAC
B-TDH3-B R	GGTCTCTCATTTTTGTTTGTTTATGTGTGTTTATTCGAAAC
MFA2-F	ggtctcatgagTTTTTGACGACAACCAAGAGGTCAAATC
MFA2-R	ggtctctAAAAAACTGGCTCAAAACTTTTTCACTTTTACG

II-4 Primers for construction of pWL032 SCRaMbLE-in pathway acceptor vector

Description	Sequence
rB-RFP-F1	CTTGATATCGAATTCCTGCAGCCCTAACTTTAAATAATTGGCATTATTTAAAGTTAcagtagagaccgaattcgcggc
rB-RFP-R1	GCTCTAGAACTAGTGGATCCCCCTAACTTTAAATAATGCCAATTATTTAAAGTTAaaaatgagaccctgcagcgg
YF compatible loxsym F	GGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCATAACTTCGTATAATGTACATTATACGAAGTTATacctagagaccccattggaattcgcg
YF compatible loxsym R	CGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCATAACTTCGTATAATGTACATTATACGAAGTTATctcatgagaccgtctgactgcag

II-5 Primers for construction of pWL121 HO integration acceptor vector

Description	Sequence
GG HO first 500 F	agcgtgGGTCTCgGGCTcctcgaggtcgacggtatcg
GG HO first 500 R	gtgctgGGTCTCaGAGTCCCGGGctcgagATGCTTTCTGAAAACACG
GG HisF	agcgtgGGTCTCgACTCTTGGCCTCCTCTAGTACACTC
GG HisR	gtgctgGGTCTCaCCTTGTCATCTTCAGTATCATACTGTTCG
GG HO up500 F	agcgtgGGTCTCgAAGGTTTAAAGTATAGATAGAATTGATTGCTGC
GG HO up500 R	gtgctgGGTCTCaCATCcaaaagctggagctccaccg
Gib RFP HO YF F	TCAGAATAGTCGTGTTTTCAGAAAGCATctcgagCCCacctagagaccccattggaattc
Gib RFP HO YF R	AAAAAAATATAGAGTGTACTAGAGGAGGCCAAGAGTCCCctcatgagaccgtctgactgc

II-6 Primers for construction of recombinase functional test related plasmids

Description	Sequence
Vox F1	GATATCGAATTCCTGCAGCCCAATAGGTCTGAGAACGCCCATTCTCAG ACGTATTcagtAGAGTGCACCATACCACAG
Vox R1	CTCTAGAACTAGTGGATCCCCCAATACGTCTGAGAATGGGCGTTCTCA GACCTATTaaaaCGGTATTTCACACCGCATAG
VloxP F1	GATATCGAATTCCTGCAGCCCTCAATTTCCGAGAATGACAGTTCTCAG AAATTGAcagtAGAGTGCACCATACCACAG
VloxP R1	CTCTAGAACTAGTGGATCCCCCTCAATTTCTGAGAACTGTCATTCTCG GAAATTGAaaaaCGGTATTTCACACCGCATAG
SloxP F1	GATATCGAATTCCTGCAGCCCCTCGTGTCCGATAACTGTAATTATCGG ACATGATcagtAGAGTGCACCATACCACAG
SloxP R1	CTCTAGAACTAGTGGATCCCCCATCATGTCCGATAATTACAGTTATCG GACACGAGaaaaCGGTATTTCACACCGCATAG
Vox/VloxP/SloxP F2	CGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCC
Vox/VloxP/SloxP R2	CCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCC
loxWT RFPYB F	ATAACTTCGTATAGCATACATTATACGAAGTTATGGCTAGAGACGgaat tcgcggccgcttctag
LoxWT RFPYB R	ATAACTTCGTATAATGTATGCTATACGAAGTTATCATCTGAGACGctgc agcggccgctactag
lox71 YB F	taccgTTCGTATAGCATACATTATACGAAGTTATGGCTAGAGACGgaatt cgcg
lox71 YB R	ATAACTTCGTATAATGTATGCTATACGAAcggtaCATCTGAGACGctgca gcg
lox66 YB F	ATAACTTCGTATAGCATACATTATACGAAcggtaGGCTAGAGACGgaatt cgcg
lox66 YB R	taccgTTCGTATAATGTATGCTATACGAAGTTATCATCTGAGACGctgcag cg
lox15 YB F	AattaTTCGTATAGCATACATTATACGAAGTTATGGCTAGAGACGgaatt cgcg
lox15 YB R	ATAACTTCGTATAATGTATGCTATACGAAtaatTCATCTGAGACGctgca gcg
lox510 YB F	taAcgTTCGTATAGCATACATTATACGAAGTTATGGCTAGAGACGgaatt cgcg
lox510 YB R	ATAACTTCGTATAATGTATGCTATACGAAcgTtaCATCTGAGACGctgca gcg
lox17 YB F	ATAACTTCGTATAGCATACATTATAgcAAtTTATGGCTAGAGACGgaatt cgcg

lox17 YB R	ATAAaTTgcTATAATGTATGCTATACGAAGTTATCATCTGAGACGctgcagcg
lox71:66 RFP F	taccgTTCGTATAGCATACATTATACGAAcggtagaattcgcggccgcttctag
lox71:17 RFP F	taccgTTCGTATAGCATACATTATAgcAAtTTATgaattcgcggccgcttctag
lox15:66 RFP F	AattaTTCGTATAGCATACATTATACGAAcggtagaattcgcggccgcttctag
lox15:17 RFP F	AattaTTCGTATAGCATACATTATAgcAAtTTATgaattcgcggccgcttctag
lox510:66 RFP F	taAcgTTCGTATAGCATACATTATACGAAcggtagaattcgcggccgcttctag
lox510:17 RFP F	taAcgTTCGTATAGCATACATTATAgcAAtTTATgaattcgcggccgcttctag
lox66 YB F2/lox17 YB F2	CCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCATAACTTCGTATAGCATACA
lox66 YB R2	CGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCtaccgTTCGTATAATGTATG
lox15 YB F2	CCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCAattaTTCGTATAGCATACA
lox510 YB F2/lox510:66 RFP F2/lox510:17 RFP F2	CCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCtaAcgTTCGTATAGCATACA
lox17 YB R2	CGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCATAAaTTgcTATAATGTATG
lox71:66 RFP F2/lox71:17 RFP F2	CCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCtaccgTTCGTATAGCATACA
loxWT Gibson RFP R2/lox15 YB R2/lox510 YB R2	CGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCATAACTTCGTATAATGTATG
lox 15 right F1	TCCCCCATAACTTCGTATAGCATACATTATACGAAtaatTGGCTAGAGACGgaattcgcg
lox 15 right R1	GAATTCCTGCAGCCCAattaTTCGTATAATGTATGCTATACGAAGTTATCATCTGAGACG
lox 15 right R2	CGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCAattaTTCG
lox510 right F1	TCCCCCATAACTTCGTATAGCATACATTATACGAAcgTtaGGCTAGAGACGgaattcgcg
lox 510 right R1	GAATTCCTGCAGCCCtaAcgTTCGTATAATGTATGCTATACGAAGTTATCATCTGAGACG
lox 510 right R2	CGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCtaAcgTTC
lox17 left F1	CTAGTGGATCCCCCATAAaTTgcTATAGCATACATTATACGAAGTTATGGCTAGAGACG
lox17 left F2	CCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCATAAaTTgcTATAG
lox17 left R1	GCAGCCCATAACTTCGTATAATGTATGCTATAgcAAtTTATCATCTGAGACGctgcagcg
lox15 LR F1	TCCCCCAattaTTCGTATAGCATACATTATACGAAtaatTgaattcgcggccgcttctag
lox510 LR F1	TCCCCCtaAcgTTCGTATAGCATACATTATACGAAcgTtaGGCTAGAGACGgaattcgcg
lox17 LR F1	AGTGGATCCCCCATAAaTTgcTATAGCATACATTATAgcAAtTTATgaattcgcggccgc
lox m3 F1	CTAGTGGATCCCCCATAACTTCGTATAtggTAttaTATACGAAGTTATGGCTAGAGACGg
lox m3/m7 F2	CCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCATAACTTCG
lox m7 F1	CTAGTGGATCCCCCATAACTTCGTATAttcTAtcTTATACGAAGTTATGGCTAGAGACGg

lox m3 R1	TTCCTGCAGCCCATAACTTCGTATAtaaTAccaTATACGAAGTTATCATCTGAGACGctg
lox m3/m7 R2	CGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCATAACTTCG
lox m7 R1	TTCCTGCAGCCCATAACTTCGTATAAgaTAgaaTATACGAAGTTATCATCTGAGACGctg
YF Vlox RFP F	agcgtgCGTCTCgGGCTTCAATTTCCGAGAATGACAGTTCTCAGAAATTGAgaattcgcggccgcttctag
YF Vlox RFP R	gtgctgCGTCTCgCATCATCAATTTCTGAGAACTGTCATTCTCGGAAATTGActgcagcggccgctactag
GG lox15 RFP F	agcgtgGGTCTCgGGCTAattaTTCGTATAGCATACATTATACGAAGTTATgaattcgcggccgcttctagagc
GG lox15 RFP R	gtgctgGGTCTCgCATCAATAACTTCGTATAATGTATGCTATACGAAtaatTctgcagcggccgctactagt
GG lox15 m3 RFP F	agcgtgGGTCTCgGGCTAattaTTCGTATAtggTAttaTATACGAAGTTATgaattcgcggccgcttctagagc
GG lox15 m3 RFP R	gtgctgGGTCTCgCATCAATAACTTCGTATAtaaTAccaTATACGAAtaatTctgcagcggccgctactagt
GG lox15 m7 RFP F	agcgtgGGTCTCgGGCTAattaTTCGTATAttcTAtcTTATACGAAGTTATgaattcgcggccgcttctagagc
GG lox15 m7 RFP R	gtgctgGGTCTCgCATCAATAACTTCGTATAAgaTAgaaTATACGAAtaatTctgcagcggccgctactagt
Gib lox17 m3 F	CGGCCGCTCTAGAACTAGTGGATCCCCCATAACTTCGTATAtggTAttaTATAgcAAtTTATGGCTAGAGACGgaattc
Gib lox17 m3 R	GATAAGCTTGATATCGAATTCCTGCAGCCCATAAaTTgcTATAtaaTAccaTATACGAAGTTATCATCTGAGACGctgc
Gib lox17 m7 F	CGGCCGCTCTAGAACTAGTGGATCCCCCATAACTTCGTATAttcTAtcTTATAgcAAtTTATGGCTAGAGACGgaattc
Gib lox17 m7 R	GATAAGCTTGATATCGAATTCCTGCAGCCCATAAaTTgcTATAAgaTAgaaTATACGAAGTTATCATCTGAGACGctgc
GG rox RFP F	agcgtgGGTCTCgGGCTtTAACTTTAAATAATTGGCATTATTTAAAGTTAgaattcgcggccgcttctagagc
GG rox RFP R	gtgctgGGTCTCgCATCAgTAACTTTAAATAATGCCAATTATTTAAAGTTActgcagcggccgctactagt
15RFP15 Unit F	agcgtgCGTCTCgGGCTAattaTTCGTATAGCATACATTATACGAAGTTATgaattcgcg
15RFP15 Unit R	gtgctgCGTCTCgCATCAATAACTTCGTATAATGTATGCTATACGAAtaatTctgcagcg
15m3RFPm3 Unit F	agcgtgCGTCTCgGGCTAattaTTCGTATAtggTAttaTATACGAAGTTATgaattcgcg
15m3RFPm3 Unit R	gtgctgCGTCTCgCATCAATAACTTCGTATAtaaTAccaTATACGAAtaatTctgcagcg
15m7RFPm7 Unit F	agcgtgCGTCTCgGGCTAattaTTCGTATAttcTAtcTTATACGAAGTTATgaattcgcg
15m7RFPm7 Unit R	gtgctgcGTCTCgCATCAATAACTTCGTATAAgaTAgaaTATACGAAtaatTctgcagcg
roxRFProx Unit F	agcgtgCGTCTCgGGCTtTAACTTTAAATAATTGGCATTATTTAAAGTTAgaattcgcgg
roxRFProx Unit R	gtgctgCGTCTCgCATCAGTAACTTTAAATAATGCCAATTATTTAAAGTTActgcagcgg
Int pro seq R	GGTCGAATCTGTAACCGTCGCG
pUC sub VURV F	CGACGGCCAGTGAATTCGAGCTCGGTACCCTCAATTTCCGAGAATGACAGTTCTCAGAAATTGAcagtAGAG

pUC sub VURV R	ATGCCTGCAGGTCGACTCTAGAGGATCCCCTCAATTTCTGAGAACTGTCATTCTCGGAAATTGACATCTGAGACG
pUC sub 17UR17 F	GGCCAGTGAATTCGAGCTCGGTACCCATAACTTCGTATAGCATACATTATAgcAAtTTATcagtAGAGTGCACCATACC
pUC sub 17UR17 R	CTGCAGGTCGACTCTAGAGGATCCCCATAAaTTgcTATAATGTATGCTATACGAAGTTATCATCTGAGACGctgcagcg
pUC sub 17m3UR17m3 F	GGCCAGTGAATTCGAGCTCGGTACCCATAACTTCGTATAtggTAttaTATAgcAAtTTATcagtAGAGTGCACCATACC
pUC sub 17m3UR17m3 R	CTGCAGGTCGACTCTAGAGGATCCCCATAAaTTgcTATAtaaTAccaTATACGAAGTTATCATCTGAGACGctgcagcg
pUC sub 17m7UR17m7 F	GGCCAGTGAATTCGAGCTCGGTACCCATAACTTCGTATAttcTAtcTTATAgcAAtTTATcagtAGAGTGCACCATACC
pUC sub 17m7UR17m7 R	CTGCAGGTCGACTCTAGAGGATCCCCATAAaTTgcTATAAgaTAgaaTATACGAAGTTATCATCTGAGACGctgcagcg
pUC sub roxURrox F	ACGGCCAGTGAATTCGAGCTCGGTACCCTAACTTTAAATAATTGGCATTATTTAAAGTTAcagtAGAGTGCACCATACC
pUC sub roxURrox R	TGCCTGCAGGTCGACTCTAGAGGATCCCCTAACTTTAAATAATGCCAATTATTTAAAGTTACATCTGAGACGctgcagc

II-7 Primers for violacein and β-carotene genes amplification

Description	Sequence
VioA-F	agcgtgGGTCTCaGATGAAGCACAGTTCTGATATATGTATTGT
VioA-R	gtgctgGGTCTCgGCTATGCCGCTATGCGC
VioB-F	agcgtgGGTCTCaGATGTCAATTTTAGACTTTCCCAGAATTCA
VioB-R	gtgctgGGTCTCgGCTATGCTTCCCTACTTAGCTTTC
VioC-F	agcgtgGGTCTCaGATGAAACGTGCAATAATCGTAGG
VioC-R	gtgctgGGTCTCgGCTAGTTGACGCGACCG
VioD-F	agcgtgGGTCTCaGATGAAGATATTGGTAATTGGTGCG
VioD-R	gtgctgGGTCTCgGCTATCTTTGCAAGGCGTATCT
VioE-F	agcgtgGGTCTCaGATGGAAAATCGTGAACCCCC
VioE-R	gtgctgGGTCTCgGCTAACGTTTGGCTGCGAACAC
CRTE-F	agcgtgGGTCTCaGATGGACTACGCTAACATCTTGAC
CRTE-R	gtgctgGGTCTCgGCTACAATGGGATGTCAGCCAAC
CRTI-F	agcgtgGGTCTCaGATGGGTAAGGAACAAGACCAAGAC
CRTI-R	gtgctgGGTCTCgGCTAGAAAGCCAAAACACCAACAGATC
CRTYB-F	agcgtgGGTCTCaGATGACCGCTTTGGCTTACTAC
CRTYB-R	gtgctgGGTCTCgGCTATTGACCTTCCCAACCAGACATAAC

II-8 Primers for promoter and terminator amplification for violacein and β-carotene TU assembly

Description	Sequence
P_{TDH3} F	agcgtgGGTCTCgGGCTgtaggtgtctcgggcagtac
P_{TDH3} R	gtgctgGGTCTCaCATCtttgtttgtttatgtgtgtttattcg
P_{PGK1} F	agcgtgGGTCTCgGGCTCTTGTCAATTCAGTTGTTTGC
P_{PGK1} R	gtgctgGGTCTCaCATCtgtTTTATATTTGTTGTAAAaaGTAG
P_{ACT1} F	agcgtgGGTCTCgGGCTttaacctacattcttccttatcgg

P_{ACT1} R	gtgctgGGTCTCaCATCtgttaattcagtaaattttcgatc
P_{RPS2} F	agcgtgGGTCTCgGGCTTTGGCTTATTCACTAAGGATTC
P_{RPS2} R	gtgctgGGTCTCaCATCtacttaattaGTTGAGCTTATTGG
P_{ZEO1} F	agcgtgGGTCTCgGGCTGTTAAACGTGTGGTTTATGG
P_{ZEO1} R	gtgctgGGTCTCaCATCtattaattgatataaACGTAGTTTTGTATG
T_{ACS2} F	agcgtgGGTCTCtTAGCCTTAAATGAGAAAAATTTCGTAaTGAG
T_{ACS2} R	gtgctgGGTCTCgGAGGATGCGAGTTCAGGTGTAACGTAG
T_{ENO2} F	agcgtgGGTCTCtTAGCagtgcttttaactaagaatTATTAG
T_{ENO2} R	gtgctgGGTCTCgGAGGTGAATGAGGGTGTACAATGC
T_{ACS1} F	agcgtgGGTCTCtTAGCgaatatgtataaagttcattTGTTC
T_{ACS1} R	gtgctgGGTCTCgGAGGCGGAGAGTTTACACCTCTTC
T_{CIT1} F	agcgtgGGTCTCtTAGCggaaaatttgatTTTTGAtTCagGG
T_{CIT1} R	gtgctgGGTCTCgGAGGGGCATGGCTTGTCGATGCAC
T_{FUM1} F	agcgtgGGTCTCtTAGCacgagctaaatacctaataatatac
T_{FUM1} R	gtgctgGGTCTCgGAGGcacgcatccaagcgtcatag

II-9 Primers for YFASS vector construction

Description	**Sequence**
Gib ccdB F	GCGGCCGGCGCGCCTCGTCATGCAGGGATacctagagaccggatccggcttactaaaagccag
Gib ccdB R	TAGCACAGAATTCTCAGGCGCGCCATCAGTTAACCCATGctcatgagacccgtcgacctgcagactggctg
YFASS1 ccdB F	GTCTGTTGTAAAAAATAATCAACACTATCCTGGTTCGGCacctagagaccggatccggcttactaaaagccag
YFASS1 ccdB R	ATCTTTTCGTCCACAGGAATCATAGTATTCAGGCGGCCTctcatgagacccgtcgacctgcagactggctg
YFASS2 ccdB F	GCACCTTACATAATACTTACGAGGACTATTATTTCTGCacctagagaccggatccggcttactaaaagccag
YFASS2 ccdB R	CATTTAAGTTACTGACATATCTTTACATCATGTCGCGGCCTctcatgagacccgtcgacctgcagactggctg
YFASS3 ccdB F	GCATTCGTCATAACCTTTAACAATGACTCTCGAGATGCacctagagaccggatccggcttactaaaagccag
YFASS3 ccdB R	TAAGTGAGTTATAGAAAATGACCCACTTACTAGCGGCCTctcatgagacccgtcgacctgcagactggctg
YFASS4 ccdB F	GTTAATCAGTCCTTGATGAACATCTGTTGTCCTAGGATGCacctagagaccggatccggcttactaaaagccag
YFASS4 ccdB R	GAAGTGTGAACCATATAAGTTCTGTAATCTTCGCGGCCTctcatgagacccgtcgacctgcagactggctg
YFASS5-24 ccdB F	GGCGAACACATCTACCATTTACCACCACAAAAGTTGAGCacctagagaccggatccggcttactaaaagccag
YFASS5-24 ccdB R	CTGGGTATAAAAAGAATAAATGTCAGAAGTTTTGCGGCCTctcatgagacccgtcgacctgcagactggctg
pRS413 BsaI acct F	GGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCacctagagaccccattggaattcg
pRS413 BsaI tgag R	CCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCctcatgagaccgtctgactgcag

II-10 Primers for Nocathiacin gene cloning

Gene name	Forward primer	Reverse primer	Reverse primer with FlAG
NocP	agcgtgGGTCTCaGATGC TGAAGATCAGAGTTCTT GGAC	gtgctgGGTCTCgGCTA GTTGCTAGCGATCGC CGGATG	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCGTTGCTAGCGATC GCCGGATG
NocS1	agcgtgGGTCTCaGATGG CTTCTCATGGTCACGTT GTTC	gtgctgGGTCTCgGCTA CGCTATCGCCACACG ACCCAA	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCCGCTATCGCCACA CGACCCAA
NocB	agcgtgGGTCTCaGATGA CTCGTGCCGACGCAGCT ACAT	gtgctgGGTCTCgGCTA CCTTGCGTCCGTTCG CTCGAA	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCCCTTGCGTCCGTTC GCTCGAA
NocQ	agcgtgGGTCTCaGATGA GCGAAACAGCCAAGAC ATCTA	gtgctgGGTCTCgGCTA GGCAGCTCTCCTTGT CATCCG	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCGGCAGCTCTCCTT GTCATCCG
NocA	agcgtgGGTCTCaGATGC ATCTGACCATTATATTA TGGG	gtgctgGGTCTCgGCTA CCTTCCATCTCCACCT TGCTT	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCCCTTCCATCTCCAC CTTGCTT
NocS2	agcgtgGGTCTCaGATGA CTTATGGTCACGCCCAC GCAG	gtgctgGGTCTCgGCTA CCCTTTTACTCCGACG AGGTA	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCCCCTTTTACTCCGA CGAGGTA
NocS3	agcgtgGGTCTCaGATGA CGCAGCATGATCGTTCG TTAA	gtgctgGGTCTCgGCTA GATTACACGGACTTC GGGGAC	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCGATTACACGGACT TCGGGGAC
NocS4	agcgtgGGTCTCaGATGT TGGGCTGCGCCGATATT GCTA	gtgctgGGTCTCgGCTA GAGGGGCGTTATAG CGGCATT	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCGAGGGGCGTTATA GCGGCATT
NocS5	agcgtgGGTCTCaGATGT TAGCTAGAGGCGCCAT TGGTG	gtgctgGGTCTCgGCTA CCCGTCCGCCCTTCCT GGTAA	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCCCCGTCCGCCCTTC CTGGTAA
NocS6	agcgtgGGTCTCaGATGT TTCAGCCATCACTGGGT GCTG	gtgctgGGTCTCgGCTA ACGCGCCACCCCGGC TGTGCG	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCACGCGCCACCCCG GCTGTGCG
NocT	agcgtgGGTCTCaGATGG CTGATACACAGGCGCC ACCAC	gtgctgGGTCTCgGCTA GCGGTAGTGCAGGG CGGTGAT	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCGCGGTAGTGCAG GGCGGTGAT
NocU	agcgtgGGTCTCaGATGA CCACCACGATTCCAGTC GAAC	gtgctgGGTCTCgGCTA GCCATCGCTCCTCGC TGGAGT	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCGCCATCGCTCCTC GCTGGAGT
NocR	agcgtgGGTCTCaGATGA CAACTCATGCGAACCAT CCTA	gtgctgGGTCTCgGCTA GGAGTCCAGCTGGAC TGGGGT	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCGGAGTCCAGCTGG ACTGGGGT
NocV	agcgtgGGTCTCaGATGT CCGTAGCTTTAGCTGCA ACAG	gtgctgGGTCTCgGCTA GCTGTAGGGCAGTGC CCTTAC	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCGCTGTAGGGCAGT GCCCTTAC
NocC	agcgtgGGTCTCaGATGA CCACCCAGGACGGTCA AGACA	gtgctgGGTCTCgGCTA TGCGATCACTGGTAG TTCCCT	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCTGCGATCACTGGT AGTTCCCT
NocD	agcgtgGGTCTCaGATGT GGAGAGCTTTGCACGT ACATC	gtgctgGGTCTCgGCTA TGGCGTGCCGATCAG TTGAGC	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCTGGCGTGCCGATC AGTTGAGC
NocE	agcgtgGGTCTCaGATGT TATTCAGAAGGGGCAC CATTG	gtgctgGGTCTCgGCTA TCTTGCATCGCAACCT CCGGT	gtgctgGGTCTCgGCTACTTATCGTCGT CATCCTTGTAATCTCTTGCATCGCAA CCTCCGGT

NocF	agcgtgGGTCTCaGATGACCGGCGATCTGGCTGAGCACA	gtgctgGGTCTCgGCTACCTTGGCATTACCGCGACTTC	gtgctgGGTCTCgGCTACTTATCGTCGTCATCCTTGTAATCCCTTGGCATTACCGCGACTTC
NocG	agcgtgGGTCTCaGATGGCAGCTCCACAGGTGGCCGGTA	gtgctgGGTCTCgGCTATGGGAATGGATGCGGAACACA	gtgctgGGTCTCgGCTACTTATCGTCGTCATCCTTGTAATCTGGGAATGGATGCGGAACACA
NocH	agcgtgGGTCTCaGATGTCAGCCGACACCCGTTGTTTCG	gtgctgGGTCTCgGCTATCTTGCACCTCCTGGAATAGC	gtgctgGGTCTCgGCTACTTATCGTCGTCATCCTTGTAATCTCTTGCACCTCCTGGAATAGC
NocI	agcgtgGGTCTCaGATGAGGTGGGGAGGGGAATCCGCCA	gtgctgGGTCTCgGCTATGACAGCTTTCCATTAGGTGA	gtgctgGGTCTCgGCTACTTATCGTCGTCATCCTTGTAATCTGACAGCTTTCCATTAGGTGA
NocK	agcgtgGGTCTCaGATGTCCAGAGAAGCAGCAGGCGACG	gtgctgGGTCTCgGCTAGGAAACGAGTGATCTTCTCAA	gtgctgGGTCTCgGCTACTTATCGTCGTCATCCTTGTAATCGGAAACGAGTGATCTTCTCAA
NocL	agcgtgGGTCTCaGATGCGCATAGAAGCTGCAACTGTAG	gtgctgGGTCTCgGCTAGCCGATCGGGATGACGGCCGC	gtgctgGGTCTCgGCTACTTATCGTCGTCATCCTTGTAATCGCCGATCGGGATGACGGCCGC
NocM	agcgtgGGTCTCaGATGAGTGCCGACCTGTCTGCGCTGA	gtgctgGGTCTCgGCTAGGACGAACAAGAGCAGGAGCA	gtgctgGGTCTCgGCTACTTATCGTCGTCATCCTTGTAATCGGACGAACAAGAGCAGGAGCA
NocN	agcgtgGGTCTCaGATGTCGACCGCTGTCTCTTTATCTT	gtgctgGGTCTCgGCTATGCCGACTCCGGTTGGCTCAT	gtgctgGGTCTCgGCTACTTATCGTCGTCATCCTTGTAATCTGCCGACTCCGGTTGGCTCAT
NocO	agcgtgGGTCTCaGATGACTTGGACGGAGCTAGTTTTTC	gtgctgGGTCTCgGCTACCCGACCGCCAGGTCTGCCAG	gtgctgGGTCTCgGCTACTTATCGTCGTCATCCTTGTAATCCCCGACCGCCAGGTCTGCCAG

II-11 Colony PCR primers in YeastFab standard Golden Gate assembly

Description	Sequence
HC_Kan seq F	GATCCTTTGATTTTCTACCG
HC_Kan seq R	CTCGATAACTCAAAAAATACG
POT colony F	CAAAAGCTGGAGCTCCACCG
POT colony R	TAGGGCGAATTGGGTACCGG

II-12 Junction PCR test primers for nocathiacin I assembly

Description	Sequence
YAC10 junc 1 F	TCTGAACCATCTTGGAAGGACCGG
YAC Noc junc 1 R	GCGAAGGCAAGCGATCAGTATACG
YAC Noc junc 2 F	gaagaggagcataggcatacattcc
YAC Noc junc 2 R	CAGCAGCAGAACGCGATTTAGCC
YAC Noc junc 3 F	CTATTGCATAATGCACTGGAAGGGG
YAC Noc junc 3 R	TGGATATACGAACCGCATCGGCC
YAC Noc junc 4 F	GGGGCAGATGTTCAAGCTATACCC
YAC Noc junc 4 R	GGTTGCTCCAGTAAGGTATTTCGC
YAC Noc junc 5 F	CTCTCTAGCAACATATTTGATCGATACC
YAC Noc junc 5 R	CCTGCAAGTTGCAACATATCCCC

YAC Noc junc 6 F	GTTTGCTGCGAAAATACATATGTACAG
YAC10 junc 6 R	GGCGTAGAGGATCAATTCCCTTG

II-13 Sequencing primers for nocathiacin I pathway on YAC10 vector

Description	**Sequence**
Nocthiacin seq 1	CCGGTGGTTATGTTTTATCCGC
Nocthiacin seq 2	CAGGCGGCTAGGTCAAGAAAAG
Nocthiacin seq 3	CGTGAATTGGCTGGTGCAGCTG
Nocthiacin seq 4	CGAGCGAACGGACGCAAGGTAG
Nocthiacin seq 5	GCGTGGTTTCCACTCCATCTTG
Nocthiacin seq 6	CTGATAGATTAGCTCCAGGTCG
Nocthiacin seq 7	CGCAGGTTGAAACTCCGTTCAG
Nocthiacin seq 8	AGGTGGAGATGGAAGGTAGCGG
Nocthiacin seq 9	GCCACCACATGGGTTATTATGG
Nocthiacin seq 10	CCAGCCGTTAATGCTGACCCTG
Nocthiacin seq 11	CGCGCGCGTACAAGCCTTACAG
Nocthiacin seq 12	GTCTTCCTGCCTTCTGCGTTTC
Nocthiacin seq 13	TCTTGGGCAACACCTCAGGCGG
Nocthiacin seq 14	CGTGTCTAGGATTTGCCATCGC
Nocthiacin seq 15	CACTGTGATCTCCAGAGCAAAG
Nocthiacin seq 16	CAACACAGTATCTCACATCGCC
Nocthiacin seq 17	ACGTCGGCATGCGATTGTTTGG
Nocthiacin seq 18	CTAGAGGCGAGATTGGCTTGCC
Nocthiacin seq 19	CTCGTCGGGCTTTACTTACGTC
Nocthiacin seq 20	GCAGAATTCGACGGTGTCCTGC
Nocthiacin seq 21	CGACCTCGTTTCTTGGATGCTG
Nocthiacin seq 22	GACAACGCGAGCCGAATGGATG
Nocthiacin seq 23	GCAACAAGATGGTGGGACTTCG
Nocthiacin seq 24	GAGGCCGCGACATGATGTAAAG
Nocthiacin seq 25	GGAGCCCGTTCATACGCTTTAC
Nocthiacin seq 26	CTATGTCGAGACAGGCCATGAG
Nocthiacin seq 27	CCTACGACAACAACCGCCACTG
Nocthiacin seq 28	CGCTTTTTAAGCTGGCATCCAG
Nocthiacin seq 29	GGTGCACCAGGGCTTTATCGTG
Nocthiacin seq 30	CGCCAAACTTGTCATCCACCTG
Nocthiacin seq 31	TACAGCTAGCGAGACAAACGCC
Nocthiacin seq 32	TGGGCAAGAGAAAATGAGCCGG
Nocthiacin seq 33	TGCCACAATTCCTGCGGGTGGC
Nocthiacin seq 34	AGTTCCGAGCGAAATAGCACCG
Nocthiacin seq 35	GGTCCTGCTGTTCTAGATCGTG

Nocthiacin seq 36	GTACCTTAGGACCGTTGAGAGG
Nocthiacin seq 37	GAGCTGCCAGAGTTGCAGAGGC
Nocthiacin seq 38	TCCACTCATTGGGAAGTCGCGG
Nocthiacin seq 39	GTCGATACAGTTCTGGTAGCAG
Nocthiacin seq 40	CCTATGCCGAGCCGGGATTTAG
Nocthiacin seq 41	GTGCCGTCCCCATGACATTTGG
Nocthiacin seq 42	AAGATGAGTGCCGACCTGTCTG
Nocthiacin seq 43	CAGCCGACACCCGTTGTTTCGC
Nocthiacin seq 44	TGGCCGAGGTCGACTTGGATGG
Nocthiacin seq 45	TGAGCCAGCTGATGTTACCCAC
Nocthiacin seq 46	CATTGGTAGCTCCACAAAGGGC
Nocthiacin seq 47	GGCGGTTACGTTGCTAGTACAG
Nocthiacin seq 48	AAGGGCAGTTTCTTTTCGCTCG
Nocthiacin seq 49	AAGCTGCGTTAGCTGCTAGATC
Nocthiacin seq 50	GCAGGGGGCGGCACAAAATAAC
Nocthiacin seq 51	TACTGCTTATGCCGGTGATCCG
Nocthiacin seq 52	CGTCCCGCTTTGCTTGAAAGAG
Nocthiacin seq 53	CCCTGCCGATAGACCACGTGTC
Nocthiacin seq 54	CCCACTCTCTTTCTCCCTTCTG
Nocthiacin seq 55	GGAGACATATGATCGCGACAGC
Nocthiacin seq 56	CCCGCTTTTTAAGCTGGCATCC
Nocthiacin seq 57	TGCACGGTCATTTCCCAGAGGG
Nocthiacin seq 58	ACCGTAGCTGAAGCTACCATTG
Nocthiacin seq 59	TGGCCTTTGAAATGGCCATGAG
Nocthiacin seq 60	AGACCCGGCCTGTTTTTACGCG
Nocthiacin seq 61	TGGCGGTCGGGTAGCGAATATG
Nocthiacin seq 63 R	CATAACTGTTCAGGACGGCGCC
Nocthiacin seq 64 R	GACCGATGAGGACACCAGGG
Nocthiacin seq 65 R	GGTACGTCATCAGCACCCGG
Noc Seq 66 R	CCAGCTGCACCACGGAGTCC
Noc seq 67 R	CTAACAACTCGTAATGCCTGCC
Noc seq 68 R	CCAACCAAGCGGGAGCACC
Noc seq 69 R	GCAACGACGTCAACAGGCCC
Noc seq 70 R	GTATGACCAGGCGGTTGGGAG
Noc seq 71 R	CAACAGGGGTGCCTGCACATCC
Noc seq 72 R	GGACCGTCCACCGGTAGGTC

II-14 Primers for RT-PCR of nocathiacin I genes

Description	Sequence
NocP FWD Set	GCA TTA GCA TGG TGG AGA G
NocS1 FWD Set	CCT GGC ACG TGG TTA TTT
NocB FWD Set	CAG CTT TGG CAG AAG TTA GA
NocQ FWD Set	CAG GCG TAG ATG TTG AGT TC
NocA FWD Set	CGC GTG GTA GAA CTG ATT G
NocS2 FWD Set	GGT CAC GTG TCT AGG ATT TG
NocS3 FWD Set	CTG ATC CGT TTG TGG TAG AG
NocS4 FWD Set	GCT ACT GGG TCA TGA TGT TAG
NocS5 FWD Set	AAG CTG ACG GAG CTA TCA
NocS6 FWD Set	TCG TAA GCT CGC CTA CTT
NocT FWD Set	CTG AAG AGG CGA TGA GAT TAG
NocU FWD Set	GCT AAT CCG GGA ACA CTA AC
NocR FWD Set	TTG AGG CGT GCA CAA TAC
NocV FWD Set	GTC GCC GAA GAA TGT ATG AG
NocC FWD Set	GTA TGC CGT CGG GTT AAT AG
NocD FWD Set	GCG TCA GAA GAG TCG AAT AC
NocE FWD Set	CAA GAC AGG CCG ACA TAA C
NocF FWD Set	GAA CCA GTC GCA GTT GAT AG
NocG FWD Set	GTT ACC GCG ACA GTT GAT AG
NocH FWD Set	GGC CCA CTG CTT TCA TTA
NocI FWD Set	GAT TTG ACG CTG AAG GTA GG
NocK FWD Set	GGG TGT TGG TGG AAG AAA
NocL FWD Set	GTG TCT GGT GTG CCT AAA G
NocM FWD Set	CTG TCT GCG CTG AAC ATC
NocN FWD Set	CTA CAC CAG GCA CGT TTA TG
NocO FWD Set	GCT ACC TGT ACA ACC AAC TG
NocP REV Set	CTC GTA ATG CCT GCC TAA TC
NocS1 REV Set	GGT CAT AAC GCT GCC TAA A
NocB REV Set	CAG GGC TAA CAG ACC ATT AC
NocQ REV Set	CAG CTA CAC CTT CTG CAT AC
NocA REV Set	TCC ATG GGT CTT CTC ATA GG
NocS2 REV Set	CTT ACC AGA TGA CCA CCA AC
NocS3 REV Set	GCT CTA GCA GTC CTT TCT TC
NocS4 REV Set	CGT AAG CGG CTC TGT AAT G
NocS5 REV Set	CAC CAC CGA TTC TTC TAA CC
NocS6 REV Set	CAG CTG AAC ACA GCC TAT C
NocT REV Set	TGC AGG GTC AGC ATT AAC
NocU REV Set	ACC TGG ATG TCT GGC TAA

NocR REV Set	CCC GTT TAC AGT ACC TTG TG
NocV REV Set	CTC TAA CTG GAG CAG TAG GT
NocC REV Set	GAC TCG TAA CCT GCC ATT AG
NocD REV Set	GTC CTC GCA ACA TCA TAC C
NocE REV Set	ATC CGC CCT AAC CAC TAA
NocF REV Set	CTT CTG CCG CAT GCA ATA
NocG REV Set	CTG TGT CAG CAT CCA GTT C
NocH REV Set	GCC AAG ACG ACC TGA TTA C
NocI REV Set	TGT ATC GGC TCT GTG GAT AG
NocK REV Set	AAC CAA GCG CAC CTA ATC
NocL REV Set	GTG CGC TAC AAG AGT TAC C
NocM REV Set	GAA CAA GAG CAG GAG CAC
NocN REV Set	CTG ACC CTG CCG AAA TAA C
NocO REV Set	CTG TGC GAT CTG GTC TAG TA
ACT1 F	CGA ACT TCC AGA TGG TCA AG
ACT1 R	GGA CAT CGA CAT CAC ACT TC

Appendix III Basic plasmids for circuit construction

Yeast plasmids

Name	Type	Marker	Source
pRS413	*Yeast Centromere plasmids (YCp)*	*HIS3*	Cai Lab
pRS415	*Yeast Centromere plasmids (YCp)*	*LEU2*	Cai Lab
pRS416	*Yeast Centromere plasmids (YCp)*	*URA3*	Cai Lab
pRS423	*Yeast Episomal plasmids (YEp)*	*HIS3*	Cai Lab

Bacteria plasmids

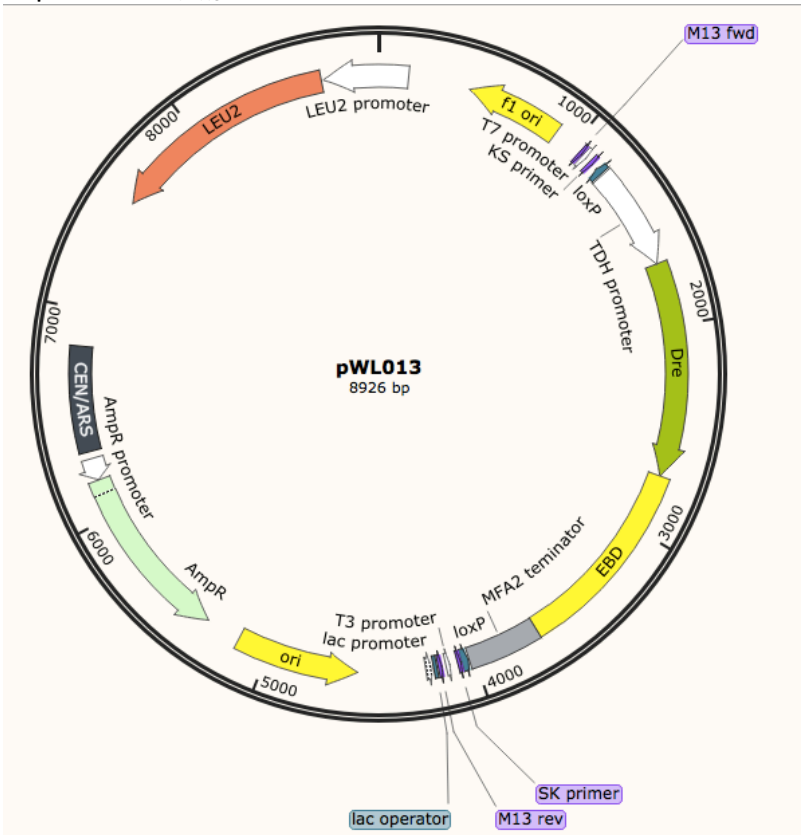
Name	Type	Marker	Source
pUC19	pUC cloning vector (high copy)	*AmpR*	NEB
pET-28 a	Bacterial Expression (high expression)	*KanR*	Jon-Marles Wright Lab

YeastFab tool plasmids

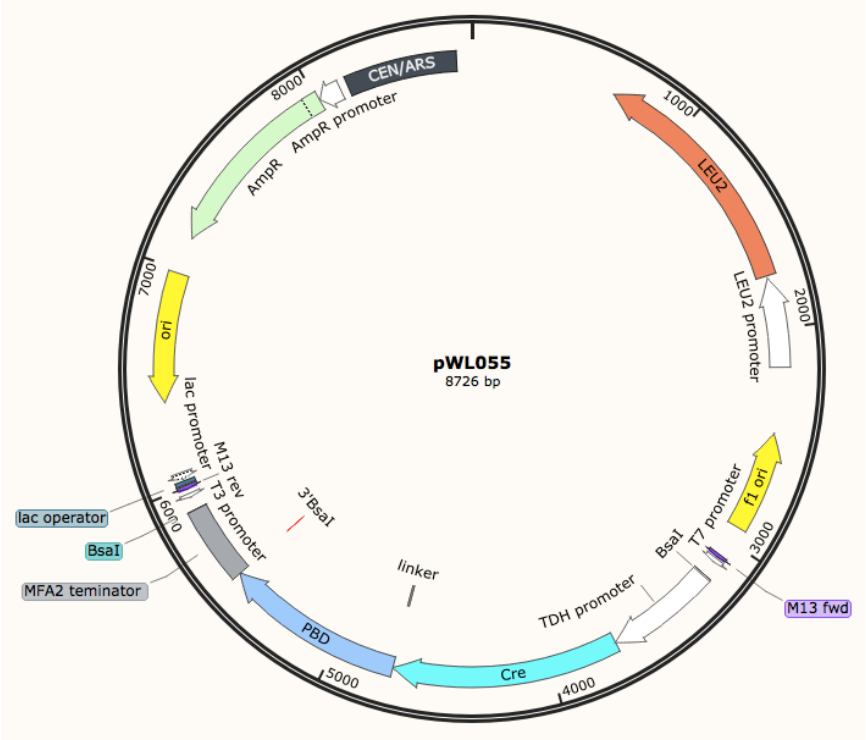
Name	**Prefix overhang of BsaI site**	**Suffix overhang of BsaI site**	**Usage**
HCKan_P	GGCT	GATG	Promoter assembly
HCKan_O	GATG	TAGC	ORF assembly
HCKan_T	TAGC	CCTC	Terminator assembly
POT1	ACCT	TGAG	1 TU assembly
POT2	ACCT	AGGC	>1 TU assembly
POT3	AGGC	TGAG	2 TUs assembly
POT4	AGGC	TGCC	>2 TUs assembly
POT5	TGCC	TGAG	3 TUs assembly
POT6	TGCC	CACT	>3 TUs assembly
POT7	CACT	TGAG	4 TUs assembly
POT8	CACT	GTCG	>4 TUs assembly
POT9	GTCG	TGAG	5 TUs assembly
POT10	GTCG	GGAG	>5 TUs assembly
POT11	GGAG	TGAG	6 TUs assembly

Appendix IV Plasmid maps of key circuits

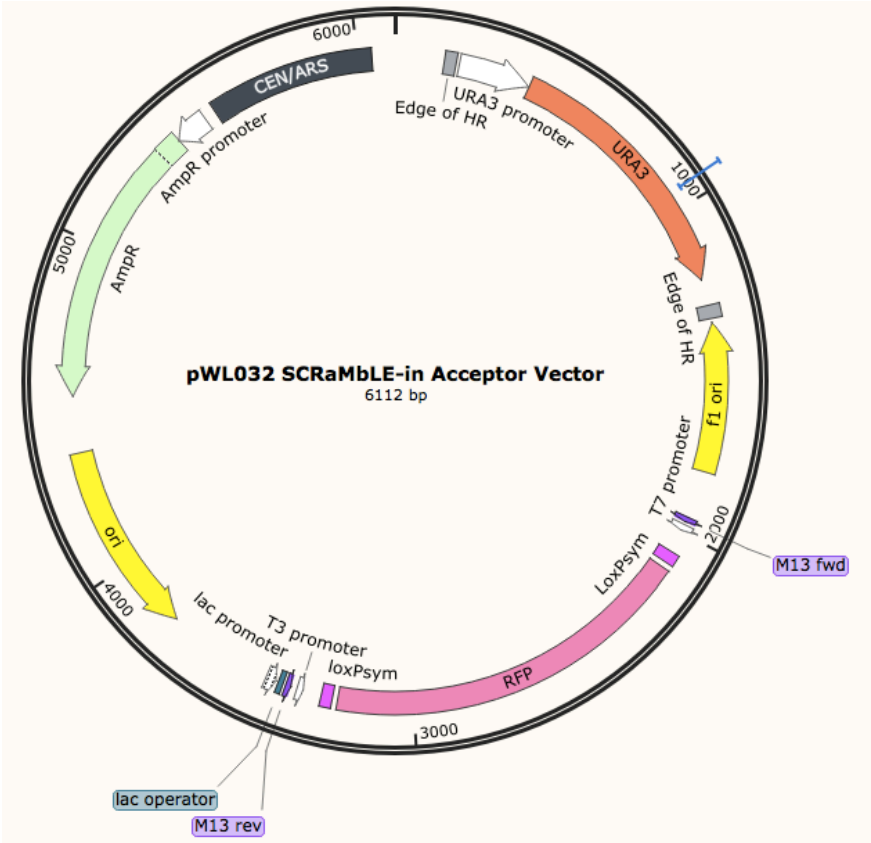
1. pWL013- P_{TDH3} DreEBD



2. pWL055- P_{TDH3} CrePBD



3. pWL032-SCRaMbLE-in acceptor vector



References:

Abil, Z., Xiong, X., & Zhao, H. (2015). Synthetic biology for therapeutic applications. *Mol Pharm, 12*(2), 322-331. doi:10.1021/mp500392q

Agapakis, C. M., Boyle, P. M., & Silver, P. A. (2012). Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nature Chemical Biology, 8*(6), 527-535. doi:10.1038/Nchembio.975

Agustin-Pavon, C., & Isalan, M. (2014). Synthetic biology and therapeutic strategies for the degenerating brain: Synthetic biology approaches can transform classical cell and gene therapies, to provide new cures for neurodegenerative diseases. *Bioessays, 36*(10), 979-990. doi:10.1002/bies.201400094

Albert, H., Dale, E. C., Lee, E., & Ow, D. W. (1995). Site-Specific Integration of DNA into Wild-Type and Mutant Lox Sites Placed in the Plant Genome. *Plant Journal, 7*(4), 649-659. doi:DOI 10.1046/j.1365-313X.1995.7040649.x

An, G. H., Bielich, J., Auerbach, R., & Johnson, E. A. (1991). Isolation and characterization of carotenoid hyperproducing mutants of yeast by flow cytometry and cell sorting. *Biotechnology (N Y), 9*(1), 70-73.

Anastassiadis, K., Fu, J., Patsch, C., Hu, S. B., Weidlich, S., Duerschke, K., . . . Stewart, A. F. (2009). Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E-coli, mammalian cells and mice. *Disease Models & Mechanisms, 2*(9-10), 508-515. doi:10.1242/dmm.003087

Anderson, J. C., Clarke, E. J., Arkin, A. P., & Voigt, C. A. (2006). Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol, 355*(4), 619-627. doi:10.1016/j.jmb.2005.10.076

Anderson, J. C., Dueber, J. E., Leguia, M., Wu, G. C., Goler, J. A., Arkin, A. P., & Keasling, J. D. (2010). BglBricks: A flexible standard for biological part assembly. *J Biol Eng, 4*(1), 1. doi:10.1186/1754-1611-4-1

Andrianantoandro, E., Basu, S., Karig, D. K., & Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol, 2*, 2006 0028. doi:10.1038/msb4100073

Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquadanio, G., Richardson, S. M., . . . Chandrasegaran, S. (2014). Total synthesis of a functional designer eukaryotic chromosome. *Science, 344*(6179), 55-58. doi:10.1126/science.1249252

Araki, K., Araki, M., & Yamamura, K. (2002). Site-directed integration of the cre gene mediated by Cre recombinase using a combination of mutant lox sites. *Nucleic Acids Res, 30*(19), e103.

Araki, K., Araki, M., & Yamamura, K. I. (1997). Targeted integration of DNA using mutant lox sites in embryonic stem cells. *Nucleic Acids Research, 25*(4), 868-872. doi:DOI 10.1093/nar/25.4.868

Araki, K., & Yamamura, K.-i. (2012). Genetic Manipulations Using Cre and Mutant LoxP Sites. In A. Morozov (Ed.), *Controlled Genetic Manipulations* (pp. 29-45). Totowa, NJ: Humana Press.

Atkinson, M. R., Savageau, M. A., Myers, J. T., & Ninfa, A. J. (2003). Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. *Cell, 113*(5), 597-607.

Atsumi, S., Hanai, T., & Liao, J. C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature, 451*(7174), 86-U13. doi:10.1038/nature06450

August, P. R., Grossman, T. H., Minor, C., Draper, M. P., MacNeil, I. A., Pemberton, J. M., . . . Osburne, M. S. (2000). Sequence analysis and functional characterization of the violacein biosynthetic pathway from Chromobacterium violaceum. *J Mol Microbiol Biotechnol, 2*(4), 513-519.

Auslander, S., & Fussenegger, M. (2013). From gene switches to mammalian designer cells: present and future prospects. *Trends Biotechnol, 31*(3), 155-168. doi:10.1016/j.tibtech.2012.11.006

Awan, A. R., Blount, B. A., Bell, D. J., Shaw, W. M., Ho, J. C. H., McKiernan, R. M., & Ellis, T. (2017). Biosynthesis of the antibiotic nonribosomal peptide penicillin in baker's yeast. *Nat Commun, 8*, 15202. doi:10.1038/ncomms15202

Ay, F., Bunnik, E. M., Varoquaux, N., Bol, S. M., Prudhomme, J., Vert, J. P., . . . Le Roch, K. G. (2014). Three-dimensional modeling of the P. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Research, 24*(6), 974-988. doi:10.1101/gr.169417.113

Baltes, N. J., & Voytas, D. F. (2015). Enabling plant synthetic biology through genome engineering. *Trends in Biotechnology, 33*(2), 120-131.

Bashor, C. J. (2008). Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics (vol 319, pg 1539, 2008). *Science, 321*(5892), 1040-1040.

Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., & Weiss, R. (2005). A synthetic multicellular system for programmed pattern formation. *Nature, 434*(7037), 1130-1134. doi:10.1038/nature03461

Basu, S., Mehreja, R., Thiberge, S., Chen, M. T., & Weiss, R. (2004). Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A, 101*(17), 6355-6360. doi:10.1073/pnas.0307571101

Baumgardner, J., Acker, K., Adefuye, O., Crowley, S. T., Deloache, W., Dickson, J. O., . . . Eckdahl, T. T. (2009). Solving a Hamiltonian Path Problem with a bacterial computer. *J Biol Eng, 3*, 11. doi:10.1186/1754-1611-3-11

Bayer, T. S., & Smolke, C. D. (2005). Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol, 23*(3), 337-343. doi:10.1038/nbt1069

Berla, B. M., Saha, R., Immethun, C. M., Maranas, C. D., Moon, T. S., & Pakrasi, H. B. (2013). Synthetic biology of cyanobacteria: unique challenges and opportunities. *Front Microbiol, 4*, 246. doi:10.3389/fmicb.2013.00246

Biggs, B. W., De Paepe, B., Santos, C. N., De Mey, M., & Kumaran Ajikumar, P. (2014). Multivariate modular metabolic engineering for pathway and strain optimization. *Curr Opin Biotechnol, 29*, 156-162. doi:10.1016/j.copbio.2014.05.005

Blake, W. J., Kaern, M., Cantor, C. R., & Collins, J. J. (2003). Noise in eukaryotic gene expression. *Nature, 422*(6932), 633-637. doi:10.1038/nature01546

Boeke, J. D., Church, G., Hessel, A., Kelley, N. J., Arkin, A., Cai, Y., . . . Yang, L. (2016). GENOME ENGINEERING. The Genome Project-Write. *Science, 353*(6295), 126-127. doi:10.1126/science.aaf6850

Bogorad, I. W., Lin, T. S., & Liao, J. C. (2013). Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature, 502*(7473), 693-697. doi:10.1038/nature12575

Bowyer, J., Jia, Z., Rosser, S., Colloms, S., & Bates, D. (2015). Development and experimental validation of a mechanistic model of in vitro DNA recombination. *Conf Proc IEEE Eng Med Biol Soc, 2015*, 945-948. doi:10.1109/EMBC.2015.7318519

Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J. C., Hieter, P., & Boeke, J. D. (1998). Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast, 14*(2), 115-132.

Brenner, K., You, L., & Arnold, F. H. (2008). Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol, 26*(9), 483-489. doi:10.1016/j.tibtech.2008.05.004

Buchholz, F., & Stewart, A. F. (2001). Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat Biotechnol, 19*(11), 1047-1052. doi:10.1038/nbt1101-1047

Cai, W., Goswami, A., Yang, Z., Liu, X., Green, K. D., Barnard-Britson, S., . . . Van Lanen, S. G. (2015). The Biosynthesis of Capuramycin-type Antibiotics: IDENTIFICATION OF THE A-102395 BIOSYNTHETIC GENE CLUSTER, MECHANISM OF SELF-RESISTANCE, AND FORMATION OF URIDINE-5'-CARBOXAMIDE. *J Biol Chem, 290*(22), 13710-13724. doi:10.1074/jbc.M115.646414

Cai, Y., Agmon, N., Choi, W. J., Ubide, A., Stracquadanio, G., Caravelli, K., . . . Boeke, J. D. (2015). Intrinsic biocontainment: multiplex genome safeguards combine transcriptional and recombinational control of essential yeast genes. *Proc Natl Acad Sci U S A, 112*(6), 1803-1808. doi:10.1073/pnas.1424704112

Cameron, D. E., Bashor, C. J., & Collins, J. J. (2014). A brief history of synthetic biology. *Nat Rev Microbiol, 12*(5), 381-390. doi:10.1038/nrmicro3239

Cantone, I., Marucci, L., Iorio, F., Ricci, M. A., Belcastro, V., Bansal, M., . . . Cosma, M. P. (2009). A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell, 137*(1), 172-181. doi:10.1016/j.cell.2009.01.055

Cantor, E. J., & Chong, S. (2001). Intein-mediated rapid purification of Cre recombinase. *Protein Expr Purif, 22*(1), 135-140. doi:10.1006/prep.2001.1428

Carothers, J. M., Goler, J. A., Juminaga, D., & Keasling, J. D. (2011). Model-Driven Engineering of RNA Devices to Quantitatively Program Gene Expression. *Science, 334*(6063), 1716-1719. doi:10.1126/science.1212209

Casini, A., MacDonald, J. T., De Jonghe, J., Christodoulou, G., Freemont, P. S., Baldwin, G. S., & Ellis, T. (2014). One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Res, 42*(1), e7. doi:10.1093/nar/gkt915

Casini, A., Storch, M., Baldwin, G. S., & Ellis, T. (2015). Bricks and blueprints: methods and standards for DNA assembly. *Nat Rev Mol Cell Biol, 16*(9), 568-576. doi:10.1038/nrm4014

Cate, D. M., Adkins, J. A., Mettakoonpitak, J., & Henry, C. S. (2015). Recent developments in paper-based microfluidic devices. *Anal Chem, 87*(1), 19-41. doi:10.1021/ac503968p

Ceroni, F., Algar, R., Stan, G. B., & Ellis, T. (2015). Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nature Methods, 12*(5), 415-418. doi:10.1038/nmeth.3339

Chae, T. U., Choi, S. Y., Kim, J. W., Ko, Y. S., & Lee, S. Y. (2017). Recent advances in systems metabolic engineering tools and strategies. *Curr Opin Biotechnol, 47*, 67-82. doi:10.1016/j.copbio.2017.06.007

Chandran, D., Bergmann, F. T., & Sauro, H. M. (2009). TinkerCell: modular CAD tool for synthetic biology. *J Biol Eng, 3*, 19. doi:10.1186/1754-1611-3-19

Chandran, D., Bergmann, F. T., Sauro, H. M., & Densmore, D. (2011). Computer-Aided Design for Synthetic Biology. *Design and Analysis of Biomolecular Circuits:Engineering Approaches to Systems and Synthetic Biology*, 203-224. doi:10.1007/978-1-4419-6766-4_10

Chang, M. C., & Keasling, J. D. (2006). Production of isoprenoid pharmaceuticals by engineered microbes. *Nat Chem Biol, 2*(12), 674-681. doi:10.1038/nchembio836

Chappell, J., Watters, K. E., Takahashi, M. K., & Lucks, J. B. (2015). A renaissance in RNA synthetic biology: new mechanisms, applications and tools for the future. *Curr Opin Chem Biol, 28*, 47-56. doi:10.1016/j.cbpa.2015.05.018

Chen, X., Zhang, C., Zou, R., Zhou, K., Stephanopoulos, G., & Too, H. P. (2013). Statistical experimental design guided optimization of a one-pot biphasic multienzyme total synthesis of amorpha-4,11-diene. *PLoS One, 8*(11), e79650. doi:10.1371/journal.pone.0079650

Chen, Y., Narendra, U., Iype, L. E., Cox, M. M., & Rice, P. A. (2000). Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol Cell, 6*(4), 885-897.

Cheng, T. H., Chang, C. R., Joy, P., Yablok, S., & Gartenberg, M. R. (2000). Controlling gene expression in yeast by inducible site-specific recombination. *Nucleic Acids Res, 28*(24), E108.

Chessher, A., Breitling, R., & Takano, E. (2015). Bacterial Microcompartments: Biomaterials for Synthetic Biology-Based Compartmentalization Strategies. *Acs Biomaterials Science & Engineering, 1*(6), 345-351. doi:10.1021/acsbiomaterials.5b00059

Choi, S. Y., Yoon, K. H., Lee, J. I., & Mitchell, R. J. (2015). Violacein: Properties and Production of a Versatile Bacterial Pigment. *Biomed Res Int, 2015*, 465056. doi:10.1155/2015/465056

Ciechonska, M., Grob, A., & Isalan, M. (2016). From noise to synthetic nucleoli: can synthetic biology achieve new insights? *Integrative Biology, 8*(4), 383-393. doi:10.1039/c5ib00271k

Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C., Osbourn, A., . . . Rosser, S. J. (2014a). Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res, 42*(4), e23. doi:10.1093/nar/gkt1101

Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C. M., Osbourn, A., . . . Rosser, S. J. (2014b). Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Research, 42*(4). doi:ARTN e23
10.1093/nar/gkt1101

Cong, L., Ran, F. A., Cox, D., Lin, S. L., Barretto, R., Habib, N., . . . Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science, 339*(6121), 819-823.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., . . . Boone, C. (2010). The genetic landscape of a cell. *Science, 327*(5964), 425-431. doi:10.1126/science.1180823

Courbet, A., Renard, E., & Molina, F. (2016). Bringing next-generation diagnostics to the clinic through synthetic biology. *Embo Molecular Medicine, 8*(9), 987-991.

d'Espaux, L., Mendez-Perez, D., Li, R., & Keasling, J. D. (2015). Synthetic biology for microbial production of lipid-based biofuels. *Curr Opin Chem Biol, 29*, 58-65. doi:10.1016/j.cbpa.2015.09.009

Dahl, R. H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T. S., Redding-Johanson, A. M., . . . Keasling, J. D. (2013). Engineering dynamic pathway regulation using stress-response promoters. *Nature Biotechnology, 31*(11), 1039-+.

Daniel, R., Rubens, J. R., Sarpeshkar, R., & Lu, T. K. (2013). Synthetic analog computation in living cells. *Nature, 497*(7451), 619-623. doi:10.1038/nature12148

Daszczuk, A., Dessalegne, Y., Drenth, I., Hendriks, E., Jo, E., van Lente, T., . . . Veening, J. W. (2014). Bacillus subtilis biosensor engineered to assess meat spoilage. *Acs Synthetic Biology, 3*(12), 999-1002. doi:10.1021/sb5000252

DeLoache, W. C., Russ, Z. N., Narcross, L., Gonzales, A. M., Martin, V. J., & Dueber, J. E. (2015). An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nat Chem Biol, 11*(7), 465-471. doi:10.1038/nchembio.1816

Dhillon, N., Raab, J., Guzzo, J., Szyjka, S. J., Gangadharan, S., Aparicio, O. M., . . . Kamakaka, R. T. (2009). DNA polymerase epsilon, acetylases and remodellers cooperate to form a specialized chromatin structure at a tRNA insulator. *EMBO J, 28*(17), 2583-2600. doi:10.1038/emboj.2009.198
tRNA 绝缘体影响蛋白 mating assay

Di Rienzi, S. C., Collingwood, D., Raghuraman, M. K., & Brewer, B. J. (2009). Fragile genomic sites are associated with origins of replication. *Genome Biol Evol, 1*, 350-363. doi:10.1093/gbe/evp034
tRNA breakpoint 基因不稳定性

Dibrov, E., Fu, S., & Lemire, B. D. (1998). The Saccharomyces cerevisiae TCM62 gene encodes a chaperone necessary for the assembly of the mitochondrial succinate dehydrogenase (complex II). *J Biol Chem, 273*(48), 32042-32048.

Ding, Y., Yu, Y., Pan, H., Guo, H., Li, Y., & Liu, W. (2010). Moving posttranslational modifications forward to biosynthesize the glycosylated thiopeptide nocathiacin I in Nocardia sp. ATCC202099. *Mol Biosyst, 6*(7), 1180-1185. doi:10.1039/c005121g

Dobrin, A., Saxena, P., & Fussenegger, M. (2016). Synthetic biology: applying biological circuits beyond novel therapies. *Integr Biol (Camb), 8*(4), 409-430. doi:10.1039/c5ib00263j

Donia, M. S., & Fischbach, M. A. (2015). HUMAN MICROBIOTA. Small molecules from the human microbiota. *Science, 349*(6246), 1254766. doi:10.1126/science.1254766

Donze, D., & Kamakaka, R. T. (2001). RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae. *EMBO J, 20*(3), 520-531. doi:10.1093/emboj/20.3.520
tRNA 抑制异染色质扩展,other elements,GAL/SAS

Du, J., Yuan, Y. B., Si, T., Lian, J. Z., & Zhao, H. M. (2012). Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Research, 40*(18). doi:10.1093/nar/gks549

Dudley, Q. M., Karim, A. S., & Jewett, M. C. (2015). Cell-free metabolic engineering: biomanufacturing beyond the cell. *Biotechnol J, 10*(1), 69-82. doi:10.1002/biot.201400330

Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., . . . Keasling, J. D. (2009). Synthetic protein scaffolds provide modular

control over metabolic flux. *Nat Biotechnol, 27*(8), 753-759. doi:10.1038/nbt.1557

Duportet, X., Wroblewska, L., Guye, P., Li, Y., Eyquem, J., Rieders, J., . . . Weiss, R. (2014). A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res, 42*(21), 13440-13451. doi:10.1093/nar/gku1082

Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., . . . Boeke, J. D. (2011). Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature, 477*(7365), 471-476. doi:10.1038/nature10403

Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature, 403*(6767), 335-338. doi:10.1038/35002125

Endy, D. (2005). Foundations for engineering biology. *Nature, 438*(7067), 449-453. doi:10.1038/nature04342

Englaender, J. A., Jones, J. A., Cress, B. F., Kuhlman, T. E., Linhardt, R. J., & Koffas, M. A. (2017). Effect of Genomic Integration Location on Heterologous Protein Expression and Metabolic Engineering in E. coli. *ACS Synth Biol*. doi:10.1021/acssynbio.6b00350

Engler, C., Gruetzner, R., Kandzia, R., & Marillonnet, S. (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One, 4*(5), e5553. doi:10.1371/journal.pone.0005553

Eroshenko, N., & Church, G. M. (2013). Mutants of Cre recombinase with improved accuracy. *Nat Commun, 4*, 2509. doi:10.1038/ncomms3509

Evans, R. M. (1988). The steroid and thyroid hormone receptor superfamily. *Science, 240*(4854), 889-895.

Farzadfard, F., & Lu, T. K. (2014). Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science, 346*(6211), 1256272. doi:10.1126/science.1256272

Flagfeldt, D. B., Siewers, V., Huang, L., & Nielsen, J. (2009). Characterization of chromosomal integration sites for heterologous gene expression in Saccharomyces cerevisiae. *Yeast, 26*(10), 545-551. doi:10.1002/yea.1705

Forster, A. C., & Church, G. M. (2007). Synthetic biology projects in vitro. *Genome Res, 17*(1), 1-6. doi:10.1101/gr.5776007

Gaj, T., Gersbach, C. A., & Barbas, C. F., 3rd. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol, 31*(7), 397-405. doi:10.1016/j.tibtech.2013.04.004

Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., & Smolke, C. D. (2015). Complete biosynthesis of opioids in yeast. *Science, 349*(6252), 1095-1100. doi:10.1126/science.aac9373

Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. A., . . . Sauro, H. M. (2014). The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat Biotechnol, 32*(6), 545-550. doi:10.1038/nbt.2891

Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. *Nature, 403*(6767), 339-342. doi:10.1038/35002131

Ghosh, K., & Van Duyne, G. D. (2002). Cre-loxP biochemistry. *Methods, 28*(3), 374-383.

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., . . . Venter, J. C. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science, 329*(5987), 52-56. doi:10.1126/science.1190719

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods, 6*(5), 343-345. doi:10.1038/nmeth.1318

Gidijala, L., Kiel, J. A., Douma, R. D., Seifar, R. M., van Gulik, W. M., Bovenberg, R. A., . . . van der Klei, I. J. (2009). An engineered yeast efficiently secreting penicillin. *PLoS One, 4*(12), e8317. doi:10.1371/journal.pone.0008317

Gietz, R. D., & Schiestl, R. H. (2007). Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nature Protocols, 2*(1), 1-4. doi:10.1038/nprot.2007.17

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., . . . Oliver, S. G. (1996). Life with 6000 genes. *Science, 274*(5287), 546, 563-547.

Gohil, V. M., Thompson, M. N., & Greenberg, M. L. (2005). Synthetic lethal interaction of the mitochondrial phosphatidylethanolamine and cardiolipin biosynthetic pathways in Saccharomyces cerevisiae. *J Biol Chem, 280*(42), 35410-35416. doi:10.1074/jbc.M505478200

Green, A. A., Silver, P. A., Collins, J. J., & Yin, P. (2014). Toehold switches: de-novo-designed regulators of gene expression. *Cell, 159*(4), 925-939. doi:10.1016/j.cell.2014.10.002

Guet, C. C., Elowitz, M. B., Hsing, W. H., & Leibler, S. (2002). Combinatorial synthesis of genetic networks. *Science, 296*(5572), 1466-1470. doi:DOI 10.1126/science.1067407

Guo, F., Gopaul, D. N., & van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature, 389*(6646), 40-46. doi:10.1038/37925

Guo, Y., Dong, J., Zhou, T., Auxillos, J., Li, T., Zhang, W., . . . Dai, J. (2015). YeastFab: the design and construction of standard biological parts for metabolic engineering in Saccharomyces cerevisiae. *Nucleic Acids Res, 43*(13), e88. doi:10.1093/nar/gkv464

Guterl, J. K., Garbe, D., Carsten, J., Steffler, F., Sommer, B., Reisse, S., . . . Sieber, V. (2012). Cell-Free Metabolic Engineering: Production of Chemicals by Minimized Reaction Cascades. *Chemsuschem, 5*(11), 2165-2172. doi:10.1002/cssc.201200365

Hagen, A., Poust, S., Rond, T., Fortman, J. L., Katz, L., Petzold, C. J., & Keasling, J. D. (2016). Engineering a Polyketide Synthase for In Vitro Production of Adipic Acid. *Acs Synthetic Biology, 5*(1), 21-27. doi:10.1021/acssynbio.5b00153

Han, D., Pal, S., Nangreave, J., Deng, Z., Liu, Y., & Yan, H. (2011). DNA origami with complex curvatures in three-dimensional space. *Science, 332*(6027), 342-346. doi:10.1126/science.1202998

Harper, A. D., Bailey, C. B., Edwards, A. D., Detelich, J. F., & Keatinge-Clay, A. T. (2012). Preparative, in vitro biocatalysis of triketide lactone chiral building blocks. *Chembiochem, 13*(15), 2200-2203. doi:10.1002/cbic.201200378

Hawkins, K. M., & Smolke, C. D. (2008). Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat Chem Biol, 4*(9), 564-573. doi:10.1038/nchembio.105

Hayashi, S., Ozaki, T., Asamizu, S., Ikeda, H., Omura, S., Oku, N., . . . Onaka, H. (2014). Genome mining reveals a minimum gene set for the biosynthesis of 32-membered macrocyclic thiopeptides lactazoles. *Chem Biol, 21*(5), 679-688. doi:10.1016/j.chembiol.2014.03.008

Haynes, K. A., Broderick, M. L., Brown, A. D., Butner, T. L., Dickson, J. O., Harden, W. L., . . . Poet, J. L. (2008). Engineering bacteria to solve the Burnt Pancake Problem. *J Biol Eng, 2*, 8. doi:10.1186/1754-1611-2-8

Hermann, M., Stillhard, P., Wildner, H., Seruggia, D., Kapp, V., Sanchez-Iranzo, H., . . . Pelczar, P. (2014). Binary recombinase systems for high-resolution conditional mutagenesis. *Nucleic Acids Res, 42*(6), 3894-3907. doi:10.1093/nar/gkt1361

Hillman, R. T., & Calos, M. P. (2012). Site-specific integration with bacteriophage PhiC31 integrase. *Cold Spring Harb Protoc, 2012*(5). doi:10.1101/pdb.prot069211

Hillson, N. J., Rosengarten, R. D., & Keasling, J. D. (2012). j5 DNA assembly design automation software. *Acs Synthetic Biology, 1*(1), 14-21. doi:10.1021/sb2000116

Hodgman, C. E., & Jewett, M. C. (2012). Cell-free synthetic biology: thinking outside the cell. *Metabolic Engineering, 14*(3), 261-269. doi:10.1016/j.ymben.2011.09.002

Hoess, R. H., Wierzbicki, A., & Abremski, K. (1986). The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res, 14*(5), 2287-2300.

Hooshangi, S., Thiberge, S., & Weiss, R. (2005). Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci U S A, 102*(10), 3581-3586. doi:10.1073/pnas.0408507102

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell, 157*(6), 1262-1278. doi:10.1016/j.cell.2014.05.010

Huang, H. H., Camsund, D., Lindblad, P., & Heidorn, T. (2010). Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Res, 38*(8), 2577-2593. doi:10.1093/nar/gkq164

Hutchison, C. A., 3rd, Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., . . . Venter, J. C. (2016). Design and synthesis of a minimal bacterial genome. *Science, 351*(6280), aad6253. doi:10.1126/science.aad6253

Isaacs, F. J., Dwyer, D. J., & Collins, J. J. (2006). RNA synthetic biology. *Nat Biotechnol, 24*(5), 545-554. doi:10.1038/nbt1208

Isaacs, F. J., Hasty, J., Cantor, C. R., & Collins, J. J. (2003). Prediction and measurement of an autoregulatory genetic module. *Proceedings of the National Academy of Sciences of the United States of America, 100*(13), 7714-7719. doi:DOI 10.1073/pnas.1332628100

Janbandhu, V. C., Moik, D., & Fassler, R. (2014). Cre recombinase induces DNA damage and tetraploidy in the absence of loxP sites. *Cell Cycle, 13*(3), 462-470. doi:10.4161/cc.27271

Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J., & Swartz, J. R. (2008). An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol Syst Biol, 4*, 220. doi:10.1038/msb.2008.57

Jones, J. A., Vernacchio, V. R., Lachance, D. M., Lebovich, M., Fu, L., Shirke, A. N., . . . Koffas, M. A. (2015). ePathOptimize: A Combinatorial Approach for Transcriptional Balancing of Metabolic Pathways. *Sci Rep, 5*, 11301. doi:10.1038/srep11301

Joshi, N., Wang, X., Montgomery, L., Elfick, A., & French, C. E. (2009). Novel approaches to biosensors for detection of arsenic in drinking water. *Desalination, 248*(1-3), 517-523.

Joung, J. K., & Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol, 14*(1), 49-55. doi:10.1038/nrm3486

Jovicevic, D., Blount, B. A., & Ellis, T. (2014). Total synthesis of a eukaryotic chromosome: Redesigning and SCRaMbLE-ing yeast. *Bioessays, 36*(9), 855-860. doi:10.1002/bies.201400086

Jullien, N., Sampieri, F., Enjalbert, A., & Herman, J. P. (2003). Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic Acids Res, 31*(21), e131.

Karimova, M., Abi-Ghanem, J., Berger, N., Surendranath, V., Pisabarro, M. T., & Buchholz, F. (2013). Vika/vox, a novel efficient and specific Cre/loxP-like site-specific recombination system. *Nucleic Acids Res, 41*(2), e37. doi:10.1093/nar/gks1037

Kaur, H., Kumar, R., Babu, J. N., & Mittal, S. (2015). Advances in arsenic biosensor development - A comprehensive review. *Biosensors & Bioelectronics, 63*, 533-545.

Ke, Y. G., Sharma, J., Liu, M. H., Jahn, K., Liu, Y., & Yan, H. (2009). Scaffolded DNA Origami of a DNA Tetrahedron Molecular Container. *Nano Letters, 9*(6), 2445-2447. doi:10.1021/nl901165f

Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metabolic Engineering, 14*(3), 189-195. doi:10.1016/j.ymben.2012.01.004

Kellendonk, C., Tronche, F., Monaghan, A. P., Angrand, P. O., Stewart, F., & Schutz, G. (1996). Regulation of Cre recombinase activity by the synthetic steroid RU 486. *Nucleic Acids Res, 24*(8), 1404-1411.

Keravala, A., Groth, A. C., Jarrahian, S., Thyagarajan, B., Hoyt, J. J., Kirby, P. J., & Calos, M. P. (2006). A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. *Mol Genet Genomics, 276*(2), 135-146. doi:10.1007/s00438-006-0129-5

Khalil, A. S., & Collins, J. J. (2010). Synthetic biology: applications come of age. *Nat Rev Genet, 11*(5), 367-379. doi:10.1038/nrg2775

Kim, J. K., Kang, H., Chae, J. S., Park, Y. H., & Choi, Y. J. (2000). Synthesis of cefminox by cell-free extracts of Streptomyces clavuligerus. *FEMS Microbiol Lett, 182*(2), 313-317.

Kis, Z., Pereira, H. S., Homma, T., Pedrigi, R. M., & Krams, R. (2015). Mammalian synthetic biology: emerging medical applications. *J R Soc Interface, 12*(106). doi:10.1098/rsif.2014.1000

Kitada, T., Kuryan, B. G., Nancy, N. H. T., Song, C., Xue, Y., Carey, M., & Grunstein, M. (2012). Mechanism for epigenetic variegation of gene expression at yeast telomeric heterochromatin. *Genes & Development, 26*(21), 2443-2455. doi:10.1101/gad.201095.112

Kosuri, S., & Church, G. M. (2014). Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods, 11*(5), 499-507. doi:10.1038/nmeth.2918

Kramer, B. P., Viretta, A. U., Daoud-El-Baba, M., Aubel, D., Weber, W., & Fussenegger, M. (2004). An engineered epigenetic transgene switch in mammalian cells. *Nat Biotechnol, 22*(7), 867-870. doi:10.1038/nbt980

Krishnakumar, R., Grose, C., Haft, D. H., Zaveri, J., Alperovich, N., Gibson, D. G., . . . Glass, J. I. (2014). Simultaneous non-contiguous deletions using large synthetic DNA and site-specific recombinases. *Nucleic Acids Research, 42*(14).

Krivoruchko, A., Siewers, V., & Nielsen, J. (2011). Opportunities for yeast metabolic engineering: Lessons from synthetic biology. *Biotechnol J, 6*(3), 262-276. doi:10.1002/biot.201000308

Kumar, R., & Thompson, E. B. (1999). The structure of the nuclear hormone receptors. *Steroids, 64*(5), 310-319.

Kuruma, Y., Stano, P., Ueda, T., & Luisi, P. L. (2009). A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells. *Biochim Biophys Acta, 1788*(2), 567-574. doi:10.1016/j.bbamem.2008.10.017

Langer, R., & Tirrell, D. A. (2004). Designing materials for biology and medicine. *Nature, 428*(6982), 487-492. doi:10.1038/nature02388

Langer, S. J., Ghafoori, A. P., Byrd, M., & Leinwand, L. (2002). A genetic screen identifies novel non-compatible loxP sites. *Nucleic Acids Res, 30*(14), 3067-3077.

Lapique, N., & Benenson, Y. (2014). Digital switching in a biosensor circuit via programmable timing of gene availability. *Nat Chem Biol, 10*(12), 1020-1027. doi:10.1038/nchembio.1680

Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J., & Dueber, J. E. (2013). Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res, 41*(22), 10668-10678. doi:10.1093/nar/gkt809

Levskaya, A., Chevalier, A. A., Tabor, J. J., Simpson, Z. B., Lavery, L. A., Levy, M., . . . Voigt, C. A. (2005). Synthetic biology: engineering Escherichia coli to see light. *Nature, 438*(7067), 441-442. doi:10.1038/nature04405

Liang, J., Ning, J. C., & Zhao, H. (2013). Coordinated induction of multi-gene pathways in Saccharomyces cerevisiae. *Nucleic Acids Res, 41*(4), e54. doi:10.1093/nar/gks1293

Lienert, F., Lohmueller, J. J., Garg, A., & Silver, P. A. (2014). Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nat Rev Mol Cell Biol, 15*(2), 95-107. doi:10.1038/nrm3738

Lindstrom, D. L., & Gottschling, D. E. (2009). The mother enrichment program: a genetic system for facile replicative life span analysis in Saccharomyces cerevisiae. *Genetics, 183*(2), 413-422, 411SI-413SI. doi:10.1534/genetics.109.106229

Linja, M. J., Porkka, K. P., Kang, Z., Savinainen, K. J., Janne, O. A., Tammela, T. L., . . . Visakorpi, T. (2004). Expression of androgen receptor coregulators in prostate cancer. *Clin Cancer Res, 10*(3), 1032-1040.

Linshiz, G., Jensen, E., Stawski, N., Bi, C. H., Elsbree, N., Jiao, H., . . . Hillson, N. J. (2016). End-to-end automated microfluidic platform for synthetic biology:

from design to functional analysis. *Journal of Biological Engineering, 10*. doi:ARTN 3
10.1186/s13036-016-0024-5

Liu, W., Yuan, J. S., & Stewart, C. N., Jr. (2013). Advanced genetic tools for plant biotechnology. *Nat Rev Genet*. doi:10.1038/nrg3626

Lloyd, A. M., & Davis, R. W. (1994). Functional expression of the yeast FLP/FRT site-specific recombination system in Nicotiana tabacum. *Mol Gen Genet, 242*(6), 653-657.

Logie, C., & Stewart, A. F. (1995). Ligand-regulated site-specific recombination. *Proc Natl Acad Sci U S A, 92*(13), 5940-5944.

Loonstra, A., Vooijs, M., Beverloo, H. B., Allak, B. A., van Drunen, E., Kanaar, R., . . . Jonkers, J. (2001). Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc Natl Acad Sci U S A, 98*(16), 9209-9214. doi:10.1073/pnas.161269798

Lynch, P. J., & Rusche, L. N. (2010). An auxiliary silencer and a boundary element maintain high levels of silencing proteins at HMR in Saccharomyces cerevisiae. *Genetics, 185*(1), 113-127. doi:10.1534/genetics.109.113100
一二级 silencer HMR-E/I tRNA conditionally maintain

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., . . . Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science, 339*(6121), 823-826. doi:10.1126/science.1232033

Mandal, M., & Breaker, R. R. (2004). Gene regulation by riboswitches. *Nat Rev Mol Cell Biol, 5*(6), 451-463. doi:10.1038/nrm1403

Mano, Y., Kobayashi, T. J., Nakayama, J., Uchida, H., & Oki, M. (2013). Single Cell Visualization of Yeast Gene Expression Shows Correlation of Epigenetic Switching between Multiple Heterochromatic Regions through Multiple Generations. *Plos Biology, 11*(7). doi:ARTN e1001601
10.1371/journal.pbio.1001601

Marcheschi, R. J., Gronenberg, L. S., & Liao, J. C. (2013). Protein engineering for metabolic engineering: current and next-generation tools. *Biotechnol J, 8*(5), 545-555. doi:10.1002/biot.201200371

Martin del Campo, J. S., Rollin, J., Myung, S., Chun, Y., Chandrayan, S., Patino, R., . . . Zhang, Y. H. P. (2013). High-Yield Production of Dihydrogen from Xylose by Using a Synthetic Enzyme Cascade in a Cell-Free System. *Angewandte Chemie-International Edition, 52*(17), 4587-4590. doi:10.1002/anie.201300766

Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., & Keasling, J. D. (2003). Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology, 21*(7), 796-802. doi:10.1038/nbt833

Mata-Gomez, L. C., Montanez, J. C., Mendez-Zavala, A., & Aguilar, C. N. (2014). Biotechnological production of carotenoids by yeasts: an overview. *Microb Cell Fact, 13*, 12. doi:10.1186/1475-2859-13-12

Maune, H. T., Han, S. P., Barish, R. D., Bockrath, M., Goddard, W. A., Rothemund, P. W. K., & Winfree, E. (2010). Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates. *Nature Nanotechnology, 5*(1), 61-66. doi:10.1038/Nnano.2009.311

McAdams, H. H., & Shapiro, L. (1995). Circuit simulation of genetic networks. *Science, 269*(5224), 650-656.

McEwan, A. R., Rowley, P. A., & Smith, M. C. (2009). DNA binding and synapsis by the large C-terminal domain of phiC31 integrase. *Nucleic Acids Res, 37*(14), 4764-4773. doi:10.1093/nar/gkp485

McIsaac, R. S., Oakes, B. L., Wang, X., Dummit, K. A., Botstein, D., & Noyes, M. B. (2013). Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res, 41*(4), e57. doi:10.1093/nar/gks1313

Medema, M. H., Breitling, R., Bovenberg, R., & Takano, E. (2011). Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. *Nat Rev Microbiol, 9*(2), 131-137. doi:10.1038/nrmicro2478

Medema, M. H., Breitling, R., & Takano, E. (2011). Synthetic Biology in Streptomyces Bacteria. *Methods in Enzymology, Vol 497: Synthetic Biology, Methods for Part/Device Characterization and Chassis Engineering, Pt A, 497*, 485-502.

Mercy, G., Mozziconacci, J., Scolari, V. F., Yang, K., Zhao, G., Thierry, A., . . . Koszul, R. (2017). 3D organization of synthetic and scrambled chromosomes. *Science, 355*(6329). doi:10.1126/science.aaf4597

Michener, J. K., Thodey, K., Liang, J. C., & Smolke, C. D. (2012). Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metabolic Engineering, 14*(3), 212-222. doi:10.1016/j.ymben.2011.09.004

Missirlis, P. I., Smailus, D. E., & Holt, R. A. (2006). A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer

region in Cre-mediated recombination. *BMC Genomics, 7*, 73. doi:10.1186/1471-2164-7-73

Mitchell, L. A., Chuang, J., Agmon, N., Khunsriraksakul, C., Phillips, N. A., Cai, Y., . . . Boeke, J. D. (2015). Versatile genetic assembly system (VEGAS) to assemble pathways for expression in S. cerevisiae. *Nucleic Acids Res, 43*(13), 6620-6630. doi:10.1093/nar/gkv466

Mitchell, L. A., Wang, A., Stracquadanio, G., Kuang, Z., Wang, X., Yang, K., . . . Boeke, J. D. (2017). Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science, 355*(6329). doi:10.1126/science.aaf4831

Miyamoto, T., Razavi, S., DeRose, R., & Inoue, T. (2013). Synthesizing biomolecule-based Boolean logic gates. *Acs Synthetic Biology, 2*(2), 72-82. doi:10.1021/sb3001112

Mootz, H. D., & Marahiel, M. A. (1997). The tyrocidine biosynthesis operon of Bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J Bacteriol, 179*(21), 6843-6850.

Muhlrad, D., & Parker, R. (1992). Mutations affecting stability and deadenylation of the yeast MFA2 transcript. *Genes Dev, 6*(11), 2100-2111.

Mularoni, L., Zhou, Y., Bowen, T., Gangadharan, S., Wheelan, S. J., & Boeke, J. D. (2012). Retrotransposon Ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in tRNA hotspots. *Genome Res, 22*(4), 693-703. doi:10.1101/gr.129460.111
0 Ty1 基因组整合位点 核小体相关性 tRNA 热点区

Nagy, A. (2000). Cre recombinase: the universal reagent for genome tailoring. *Genesis, 26*(2), 99-109.

Nern, A., Pfeiffer, B. D., Svoboda, K., & Rubin, G. M. (2011). Multiple new site-specific recombinases for use in manipulating animal genomes. *Proceedings of the National Academy of Sciences of the United States of America, 108*(34), 14198-14203.

Niederholtmeyer, H., Sun, Z. Z., Hori, Y., Yeung, E., Verpoorte, A., Murray, R. M., & Maerkl, S. J. (2015). Rapid cell-free forward engineering of novel genetic ring oscillators. *Elife, 4*, e09771. doi:10.7554/eLife.09771

Nielsen, J., & Keasling, J. D. (2011). Synergies between synthetic biology and metabolic engineering. *Nature Biotechnology, 29*(8), 693-695. doi:10.1038/nbt.1937

Nolan, E. M., & Walsh, C. T. (2009). How Nature Morphs Peptide Scaffolds into Antibiotics. *Chembiochem, 10*(1), 34-53. doi:10.1002/cbic.200800438

Olorunniji, F. J., McPherson, A. L., Rosser, S. J., Smith, M. C. M., Colloms, S. D., & Stark, W. M. (2017). Control of serine integrase recombination directionality by fusion with the directionality factor. *Nucleic Acids Res*. doi:10.1093/nar/gkx567

Paddon, C. J., & Keasling, J. D. (2014a). Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nature Reviews Microbiology, 12*(5), 355-367. doi:10.1038/nrmicro3240

Paddon, C. J., & Keasling, J. D. (2014b). Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat Rev Microbiol, 12*(5), 355-367. doi:10.1038/nrmicro3240

Pardee, K., Green, A. A., Ferrante, T., Cameron, D. E., DaleyKeyser, A., Yin, P., & Collins, J. J. (2014). Paper-based synthetic gene networks. *Cell, 159*(4), 940-954. doi:10.1016/j.cell.2014.10.004

Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., . . . Collins, J. J. (2016). Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell, 165*(5), 1255-1266. doi:10.1016/j.cell.2016.04.059

Park, S. R., Yoo, Y. J., Ban, Y. H., & Yoon, Y. J. (2010). Biosynthesis of rapamycin and its regulation: past achievements and recent progress. *J Antibiot (Tokyo), 63*(8), 434-441. doi:10.1038/ja.2010.71

Pucci, M. J., Bronson, J. J., Barrett, J. F., DenBleyker, K. L., Discotto, L. F., Fung-Tomc, J. C., & Ueda, Y. (2004). Antimicrobial evaluation of nocathiacins, a thiazole peptide class of antibiotics. *Antimicrob Agents Chemother, 48*(10), 3697-3701. doi:10.1128/AAC.48.10.3697-3701.2004

Quan, J., & Tian, J. (2009). Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One, 4*(7), e6441. doi:10.1371/journal.pone.0006441

Quinn, J. Y., Cox, R. S., 3rd, Adler, A., Beal, J., Bhatia, S., Cai, Y., . . . Sauro, H. M. (2015). SBOL Visual: A Graphical Language for Genetic Designs. *PLoS Biol, 13*(12), e1002310. doi:10.1371/journal.pbio.1002310

Quintero, M. J., Maya, D., Arevalo-Rodriguez, M., Cebolla, A., & Chavez, S. (2007). An improved system for estradiol-dependent regulation of gene expression in yeast. *Microb Cell Fact, 6*, 10. doi:10.1186/1475-2859-6-10

Redden, H., & Alper, H. S. (2015). The development and characterization of synthetic minimal yeast promoters. *Nat Commun, 6*, 7810. doi:10.1038/ncomms8810

Richardson, S. M., Mitchell, L. A., Stracquadanio, G., Yang, K., Dymond, J. S., DiCarlo, J. E., . . . Bader, J. S. (2017). Design of a synthetic yeast genome. *Science, 355*(6329), 1040-1044. doi:10.1126/science.aaf4557

Richardson, S. M., Wheelan, S. J., Yarrington, R. M., & Boeke, J. D. (2006). GeneDesign: rapid, automated design of multikilobase synthetic genes. *Genome Res, 16*(4), 550-556. doi:10.1101/gr.4431306

Rodrigues, A. L., Trachtmann, N., Becker, J., Lohanatha, A. F., Blotenberg, J., Bolten, C. J., . . . Wittmann, C. (2013). Systems metabolic engineering of Escherichia coli for production of the antitumor drugs violacein and deoxyviolacein. *Metabolic Engineering, 20*, 29-41. doi:10.1016/j.ymben.2013.08.004

Ruder, W. C., Lu, T., & Collins, J. J. (2011). Synthetic Biology Moving into the Clinic. *Science, 333*(6047), 1248-1252.

Rutherford, K., Yuan, P., Perry, K., Sharp, R., & Van Duyne, G. D. (2013). Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res, 41*(17), 8341-8356. doi:10.1093/nar/gkt580

Samodelov, S. L., Beyer, H. M., Guo, X., Augustin, M., Jia, K. P., Baz, L., . . . Zurbriggen, M. D. (2016). StrigoQuant: A genetically encoded biosensor for quantifying strigolactone activity and specificity. *Sci Adv, 2*(11), e1601266. doi:10.1126/sciadv.1601266

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., . . . Orzaez, D. (2013). GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol, 162*(3), 1618-1631. doi:10.1104/pp.113.217661

Sauer, B. (1987). Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. *Mol Cell Biol, 7*(6), 2087-2096.

Sauer, B., & McDermott, J. (2004). DNA recombination with a heterospecific Cre homolog identified from comparison of the pac-c1 regions of P1-related phages. *Nucleic Acids Res, 32*(20), 6086-6095. doi:10.1093/nar/gkh941

Schmidt, D., & Cho, Y. K. (2015). Natural photoreceptors and their application to synthetic biology. *Trends in Biotechnology, 33*(2), 80-91. doi:10.1016/j.tibtech.2014.10.007

Schwenk, F., Kuhn, R., Angrand, P. O., Rajewsky, K., & Stewart, A. F. (1998). Temporally and spatially regulated somatic mutagenesis in mice. *Nucleic Acids Res, 26*(6), 1427-1432.

Serganov, A., & Patel, D. J. (2007). Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet, 8*(10), 776-790. doi:10.1038/nrg2172

Shao, Z., Zhao, H., & Zhao, H. (2009). DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res, 37*(2), e16. doi:10.1093/nar/gkn991

Shen, Y., Stracquadanio, G., Wang, Y., Yang, K., Mitchell, L. A., Xue, Y., . . . Bader, J. S. (2016). SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Res, 26*(1), 36-49. doi:10.1101/gr.193433.115

Shen, Y., Wang, Y., Chen, T., Gao, F., Gong, J., Abramczyk, D., . . . Yang, H. (2017). Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science, 355*(6329). doi:10.1126/science.aaf4791

Sheren, J., Langer, S. J., & Leinwand, L. A. (2007). A randomized library approach to identifying functional lox site domains for the Cre recombinase. *Nucleic Acids Res, 35*(16), 5464-5473. doi:10.1093/nar/gkm604

Shetty, R. P., Endy, D., & Knight, T. F., Jr. (2008). Engineering BioBrick vectors from BioBrick parts. *J Biol Eng, 2*, 5. doi:10.1186/1754-1611-2-5

Siddiqui, M. S., Thodey, K., Trenchard, I., & Smolke, C. D. (2012). Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res, 12*(2), 144-170. doi:10.1111/j.1567-1364.2011.00774.x

Singh, S., Ghosh, P., & Hatfull, G. F. (2013). Attachment site selection and identity in Bxb1 serine integrase-mediated site-specific recombination. *PLoS Genet, 9*(5), e1003490. doi:10.1371/journal.pgen.1003490

Siuti, P., Yazbek, J., & Lu, T. K. (2013). Synthetic circuits integrating logic and memory in living cells. *Nat Biotechnol, 31*(5), 448-452. doi:10.1038/nbt.2510

Slomovic, S., Pardee, K., & Collins, J. J. (2015). Synthetic biology devices for in vitro and in vivo diagnostics. *Proc Natl Acad Sci U S A, 112*(47), 14429-14435. doi:10.1073/pnas.1508521112

Smanski, M. J., Bhatia, S., Zhao, D., Park, Y., L, B. A. W., Giannoukos, G., . . . Voigt, C. A. (2014). Functional optimization of gene clusters by combinatorial

design and assembly. *Nat Biotechnol, 32*(12), 1241-1249. doi:10.1038/nbt.3063

Smanski, M. J., Zhou, H., Claesen, J., Shen, B., Fischbach, M. A., & Voigt, C. A. (2016). Synthetic biology to access and expand nature's chemical diversity. *Nat Rev Microbiol, 14*(3), 135-149. doi:10.1038/nrmicro.2015.24

Smith, M. C., Brown, W. R., McEwan, A. R., & Rowley, P. A. (2010). Site-specific recombination by phiC31 integrase and other large serine recombinases. *Biochem Soc Trans, 38*(2), 388-394. doi:10.1042/BST0380388

Smolke, C. D. (2009). Building outside of the box: iGEM and the BioBricks Foundation. *Nat Biotechnol, 27*(12), 1099-1102. doi:10.1038/nbt1209-1099

Stark, W. M. (2017). Making serine integrases work for us. *Curr Opin Microbiol, 38*, 130-136. doi:10.1016/j.mib.2017.04.006

Stephanopoulos, G. (2012). Synthetic biology and metabolic engineering. *ACS Synth Biol, 1*(11), 514-525. doi:10.1021/sb300094q

Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., & Hasty, J. (2008). A fast, robust and tunable synthetic gene oscillator. *Nature, 456*(7221), 516-U539. doi:10.1038/nature07389

Suzuki, E., & Nakayama, M. (2011). VCre/VloxP and SCre/SloxP: new site-specific recombination systems for genome engineering. *Nucleic Acids Res, 39*(8), e49. doi:10.1093/nar/gkq1280

Tahlan, K., Ahn, S. K., Sing, A., Bodnaruk, T. D., Willems, A. R., Davidson, A. R., & Nodwell, J. R. (2007). Initiation of actinorhodin export in Streptomyces coelicolor. *Mol Microbiol, 63*(4), 951-961.

Thodey, K., Galanie, S., & Smolke, C. D. (2014). A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat Chem Biol, 10*(10), 837-844. doi:10.1038/nchembio.1613

Thomson, J. G., Rucker, E. B., 3rd, & Piedrahita, J. A. (2003a). Mutational analysis of loxP sites for efficient Cre-mediated insertion into genomic DNA. *Genesis, 36*(3), 162-167. doi:10.1002/gene.10211

Thomson, J. G., Rucker, E. B., & Piedrahita, J. A. (2003b). Mutational analysis of LoxP sites for efficient Cre-mediated insertion into genomic DNA. *Genesis, 36*(3), 162-167.

Thorpe, H. M., & Smith, M. C. (1998). In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci U S A, 95*(10), 5505-5510.

Tigges, M., Marquez-Lago, T. T., Stelling, J., & Fussenegger, M. (2009). A tunable synthetic mammalian oscillator. *Nature, 457*(7227), 309-312. doi:10.1038/nature07616

Trubitsyna, M., Michlewski, G., Cai, Y., Elfick, A., & French, C. E. (2014). PaperClip: rapid multi-part DNA assembly from existing libraries. *Nucleic Acids Res, 42*(20), e154. doi:10.1093/nar/gku829

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., & Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nat Rev Genet, 11*(9), 636-646. doi:10.1038/nrg2842

Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *P T, 40*(4), 277-283.

Verwaal, R., Wang, J., Meijnen, J. P., Visser, H., Sandmann, G., van den Berg, J. A., & van Ooyen, A. J. (2007). High-level production of beta-carotene in Saccharomyces cerevisiae by successive transformation with carotenogenic genes from Xanthophyllomyces dendrorhous. *Appl Environ Microbiol, 73*(13), 4342-4350. doi:10.1128/AEM.02759-06

Wagner, H. J., Sprenger, A., Rebmann, B., & Weber, W. (2016). Upgrading biomaterials with synthetic biological modules for advanced medical applications. *Adv Drug Deliv Rev, 105*(Pt A), 77-95. doi:10.1016/j.addr.2016.05.004

Wang, B., Wang, J., Zhang, W., & Meldrum, D. R. (2012). Application of synthetic biology in cyanobacteria and algae. *Front Microbiol, 3*, 344. doi:10.3389/fmicb.2012.00344

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature, 460*(7257), 894-898. doi:10.1038/nature08187

Wang, L., Qian, K., Huang, Y., Jin, N., Lai, H., Zhang, T., . . . Wang, D. (2015). SynBioLGDB: a resource for experimentally validated logic gates in synthetic biology. *Sci Rep, 5*, 8090. doi:10.1038/srep08090

Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A., & Gill, R. T. (2010). Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nature Biotechnology, 28*(8), 856-U138. doi:10.1038/nbt.1653

Watstein, D. M., McNerney, M. P., & Styczynski, M. P. (2015). Precise metabolic engineering of carotenoid biosynthesis in Escherichia coli towards a low-cost biosensor. *Metabolic Engineering, 31*, 171-180. doi:10.1016/j.ymben.2015.06.007

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS One, 6*(2), e16765. doi:10.1371/journal.pone.0016765

Weber, W., & Fussenegger, M. (2011). Emerging biomedical applications of synthetic biology. *Nat Rev Genet, 13*(1), 21-35. doi:10.1038/nrg3094

Weber, W., Stelling, J., Rimann, M., Keller, B., Daoud-El Baba, M., Weber, C. C., . . . Fussenegger, M. (2007). A synthetic time-delay circuit in mammalian cells and mice. *Proceedings of the National Academy of Sciences of the United States of America, 104*(8), 2643-2648. doi:10.1073/pnas.0606398104

Wood, A. J., Lo, T. W., Zeitler, B., Pickle, C. S., Ralston, E. J., Lee, A. H., . . . Meyer, B. J. (2011). Targeted Genome Editing Across Species Using ZFNs and TALENs. *Science, 333*(6040), 307-307. doi:10.1126/science.1207773

Wright, O., Stan, G. B., & Ellis, T. (2013). Building-in biosafety for synthetic biology. *Microbiology, 159*(Pt 7), 1221-1235. doi:10.1099/mic.0.066308-0

Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., & Koffas, M. A. (2016). Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol, 34*(8), 652-664. doi:10.1016/j.tibtech.2016.02.010

Wu, Y., Li, B. Z., Zhao, M., Mitchell, L. A., Xie, Z. X., Lin, Q. H., . . . Yuan, Y. J. (2017). Bug mapping and fitness testing of chemically synthesized chromosome X. *Science, 355*(6329). doi:10.1126/science.aaf4706

Wunderlich, F. T., Wildner, H., Rajewsky, K., & Edenhofer, F. (2001). New variants of inducible Cre recombinase: a novel mutant of Cre-PR fusion protein exhibits enhanced sensitivity and an expanded range of inducibility. *Nucleic Acids Res, 29*(10), E47.

Xie, Z. X., Li, B. Z., Mitchell, L. A., Wu, Y., Qi, X., Jin, Z., . . . Yuan, Y. J. (2017). "Perfect" designer chromosome V and behavior of a ring derivative. *Science, 355*(6329). doi:10.1126/science.aaf4704

Xu, Z., Thomas, L., Davies, B., Chalmers, R., Smith, M., & Brown, W. (2013). Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol, 13*, 87. doi:10.1186/1472-6750-13-87

Yadav, V. G., De Mey, M., Lim, C. G., Ajikumar, P. K., & Stephanopoulos, G. (2012). The future of metabolic engineering and synthetic biology: towards a systematic practice. *Metab Eng, 14*(3), 233-241.

Yamanishi, M., Ito, Y., Kintaka, R., Imamura, C., Katahira, S., Ikeuchi, A., . . . Matsuyama, T. (2013). A Genome-Wide Activity Assessment of Terminator Regions in Saccharomyces cerevisiae Provides a "Terminatome" Toolbox. *Acs Synthetic Biology, 2*(6), 337-347. doi:10.1021/sb300116y

Yetisen, A. K., Akram, M. S., & Lowe, C. R. (2013). Paper-based microfluidic point-of-care diagnostic devices. *Lab Chip, 13*(12), 2210-2251. doi:10.1039/c3lc50169h

Yokobayashi, Y., Weiss, R., & Arnold, F. H. (2002). Directed evolution of a genetic circuit. *Proc Natl Acad Sci U S A, 99*(26), 16587-16591. doi:10.1073/pnas.252535999

Yoshikawa, K., Furusawa, C., Hirasawa, T., & Shimizu, H. (2008). Genome-wide analysis of the effects of location and number of stress response elements on gene expression in Saccharomyces cerevisiae. *J Biosci Bioeng, 106*(5), 507-510. doi:10.1263/jbb.106.507

You, L., Cox, R. S., 3rd, Weiss, R., & Arnold, F. H. (2004). Programmed population control by cell-cell communication and regulated killing. *Nature, 428*(6985), 868-871. doi:10.1038/nature02491

Zhang, L., Zhao, G., & Ding, X. (2011). Tandem assembly of the epothilone biosynthetic gene cluster by in vitro site-specific recombination. *Sci Rep, 1*, 141. doi:10.1038/srep00141

Zhang, W., Zhao, G., Luo, Z., Lin, Y., Wang, L., Guo, Y., . . . Dai, J. (2017). Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science, 355*(6329). doi:10.1126/science.aaf3981

Zhang, Y., Werling, U., & Edelmann, W. (2012). SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res, 40*(8), e55. doi:10.1093/nar/gkr1288

Zhou, K., Qiao, K. J., Edgar, S., & Stephanopoulos, G. (2015). Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nature Biotechnology, 33*(4), 377-U157. doi:10.1038/nbt.3095

Zhu, F., Zhong, X., Hu, M., Lu, L., Deng, Z., & Liu, T. (2014). In vitro reconstitution of mevalonate pathway and targeted engineering of farnesene

overproduction in Escherichia coli. *Biotechnol Bioeng, 111*(7), 1396-1405. doi:10.1002/bit.25198

Zumla, A., Nahid, P., & Cole, S. T. (2013). Advances in the development of new tuberculosis drugs and treatment regimens. *Nat Rev Drug Discov, 12*(5), 388-404. doi:10.1038/nrd4001